Universidade de Aveiro
Ano 2013

Departamento de Ciências Sociais, Políticas e do Território

MIGUEL LOPES BATISTA VIEGAS

ÍNDICES DE PREÇOS IMOBILIÁRIOS: UM EXERCÍCIO NA ÁREA AVEIRO-ÍLHAVO

RPPI CONSTRUCTION: AN EMPIRICAL EXERCISE ON AVEIRO-ÍLHAVO URBAN AREA

Universidade de Aveiro
Ano 2013

Departamento de Ciências Sociais, Políticas e do Território

MIGUEL LOPES BATISTA VIEGAS

ÍNDICES DE PREÇOS IMOBILIÁRIOS: UM EXERCÍCIO NA ÁREA AVEIRO-ÍLHAVO

RPPI CONSTRUCTION: AN EMPIRICAL EXERCISE ON AVEIRO-ÍLHAVO URBAN AREA

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Planeamento Regional e Urbano realizada sob a orientação científica do Doutor João José Lourenço Marques, professor auxiliar do Departamento de Ciências Sociais, Políticas e do Território da Universidade de Aveiro.

O júri

Presidente:	Doutor Paulo António dos Santos Silva Professor Auxiliar do Departamento de Ciências Sociais, Políticas e do Território da Universidade de Aveiro.
Vogais:	Doutor Joaquim Carlos da Costa Pinho Professor Associado do Departamento de Economia, Gestão e Engenharia Industrial da Universidade de Aveiro (arguente)
	Doutor João José Lourenço Marques Professor Auxiliar do Departamento de Ciências Sociais, Políticas e do Território da Universidade de Aveiro (orientador)

Agradecimentos

Este trabalho foi realizado no âmbito do projecto de investigação "Factores determinantes da procura da habitação em Portugal" – DONUT (PTDC /AURURB / 100592/2008). Assim, agradeço a toda a equipa do projecto e ainda o apoio da GOVCOPP – Unidade de Investigação em Governança, Competitividade e Políticas Públicas da Universidade de Aveiro, da Fundação para a Ciência e Tecnologia (FCT), do Programa Operacional Temático Factores de Competitividade (COMPETE) do Quadro Comunitário de Apoio (QCA IV - Comissão Europeia) e do Fundo Comunitário Europeu (FEDER).

palavras-chave Índice de preços, imobiliário, dependência espacial

resumo

O mercado imobiliário ocupa um lugar central dentro de qualquer sociedade. A propriedade imobiliária representa a parte essencial da riqueza da maioria das famílias, bem como uma parte substancial das suas despesas mensais. Ele também representa uma grande fracção do investimento privado, gerando ganhos significativos. No entanto, e apesar desta importância, a compilação de índices de preços imobiliários está longe de ser satisfatória em Portugal, revelando, como veremos, uma importante lacuna no domínio da informação estatística.

A construção de um índice de preços imobiliário coloca problemas específicos decorrentes da própria natureza do objecto em questão. Considerando a carência em Portugal relativamente à falta de uma base de dados suficientemente alargada que permita a compilação de índices de preços confiáveis e capazes de melhorar o funcionamento dos mercados de habitação, esta tese procura dar um contributo nesta área, propondo uma metodologia aplicada para a área urbana de Aveiro e Ílhavo. Além disso, numa extensão final do modelo, medimos a dependência espacial através de uma análise de dados georreferenciados, procurando assim inferir o grau em que os estimadores calculados a partir de um modelo de dependência espacial contribuem para aumentar a robustez dos índices de preços.

keywords Residential property Indexes, dwelling, spatial dependence.

abstract

The housing market occupies a central place within any society. Residential property represents the essential parts of most families' wealth, as well as a substantial part of their monthly expenditure. It also represents a large fraction of private investment, generating significant earnings. However and despite this importance, the compilation of reliable residential property price indexes indices (RPPI) is far from satisfactory in Portugal, revealing, as we shall see, an important gap in the field of statistical information.

The construction of a residential property price index poses particular problems arising from the inherent nature of the object concerned. Considering the gap in Portugal regarding the lack of a sufficiently enlarged database allowing the compilation of reliable RPPIs designated to improve the operation of housing markets, this thesis seeks to make a contribution in this area by proposing a methodology applied to the urban area of Aveiro and Ílhavo. Moreover, in a final extension of the model, we measure the spatial dependence by a spatial data analysis, thus seeking to infer the extent to which the estimators calculated from a model of spatial dependence contribute to increasing the robustness of our price indices.

Table of Contents

1. Introduction

The housing market occupies a central place within any society. Residential property represents the essential parts of most families' wealth, as well as a substantial part of their monthly expenditure. It also represents a large fraction of private investment, generating significant earnings. The construction sector in Portugal currently employs about 10% of workforce, and is also responsible for about half of Gross Fixed Capital Formation (INE, 2011). Apart from its importance as a sector individualized, housing construction generates significant multiplier effects in several sectors upstream and downstream. For the amount side, we may be cited as an example the suppliers of materials (cement, steel, glass, paints, etc.), products (wood, electrical and HVAC equipment, etc.) and services (consulting, architecture, engineering, etc.) and for the downstream side, companies supplying products (appliances, furniture, etc.) and services (energy, maintenance, etc.), constituting one of the most complex chains of interdependencies. Furthermore, housing is a commodity whose access must be provided in many cases by the state (more or less supplementary), leading the authorities to intervene in a more or less incisive way in this area, either directly, either through incentives or through the soil policy. Considering the decisive role and influence of housing and real estate on the behavior of the global economy, it seems crucial that agents should have reliable information about housing prices and its evolution, since these determine the directly book value of assets, the composition of the household budget as well as the profitability of investments in real estate. As it is known, the information and transparency are key attributes for the proper functioning of the markets, and the housing market is certainly not an exception to this rule. However and despite this importance, the compilation of reliable residential property price indexes indices (RPPI) is far from satisfactory in Portugal, revealing, as we shall see, an important gap in the field of statistical information.

The construction of a residential property price index poses particular problems arising from the inherent nature of the object concerned. The fact that each dwelling constitutes a very unique and not reproducible good, combined with its durability prevents the conceptualization of a fixed basket of goods whose prices are recorded over time. Indeed, in the case of housing, prices are rarely observed in view of the sporadic nature of transactions. Their durability makes the quality varies over time due to the depreciation or due to improvements and renovations works. Considering these difficulties, it is not possible to construct a perfect index of housing prices. We can only aspire to build an approximation as accurate as possible to the theoretical RPPI ideally built (Baumont, 2009; Eurostat, 2011). The literature describes four main methods of constructing RPPIs: stratification or mix adjustment, repeat sales methods, hedonic regression methods and the use of property assessment information or appraisal-based methods (Eurostat, 2011). The proper choice depends on the available database, and also on the type of targets that we want to achieve. When the volume of information is large, allowing for sufficient homogeneous grouping of observations the stratification or mix adjustment may be the most appropriate method. It only requires information about price and location in order to calculate central price tendency estimations such as mean and median prices. This method is among the less data intensive methods and has also the advantage of being easily understood by the agents (Mark & Goldberg, 1984; McDonald & Smith, 2009). The hedonic regression methods is generally regarded as the most appropriate method for constructing price indices for housing (Eurostat, 2011). Counting that all relevant variables are included, this method maximizes the use of available information and adjusts the RPPI to both sampling variation and qualitative changes of housing, be that a consequence of depreciation or renovation (R. J. Hill, 2004; R. J. Hill & Melser, 2008). The repeat sales method consists in observing the evolution of the price of a specific housing over time. The scarcity of data naturally raised strong barriers to their implementation often leading to biased results (Shiller, 1991; Wang & Zorn, 1997). The appraisal-based method becomes attractive in countries where the government conducts systematic reviews of all properties for tax purposes. Unfortunately it does not permit incorporating qualitative changes (depreciation or renovation) and always raises important questions about the method and the reliability of the appraisal methodology.

Considering the gap in Portugal regarding the lack of a sufficiently enlarged database allowing the compilation of reliable RPPIs designated to improve the operation of housing markets, this thesis seeks to make a contribution in this area by testing several methodologies applied to the urban area of Aveiro and Ílhavo. Using the hedonic price approach, in its various forms (time dummy variables, imputation hedonic price and stratification methods), it proposes a model of compromise between completeness and parsimony that can be replicated in other localities, covering the various sub-markets, heterogeneity and geographical features and key attributes. Although hedonic models can incorporate effects of neighborhood or proximity, the literature recognizes the complexity of these effects and the difficulty in finding variables that are representative and capture these spatial effects (Basu & Thibodeau, 1998). That is, even in models including neighborhood and accessibility variables, the residual produced by these hedonic models may exhibit patterns of autocorrelation by misspecification of the model, creating inconsistency and inefficiency in OLS estimators (Anselin, 1988). Thus, in a final extension of our model, we measure the spatial dependence by a spatial data analysis, thus seeking to infer the extent to which the estimators calculated from a model of spatial dependence contribute to increasing the robustness of our price index.

After the introduction, the thesis continues with a section designed to describe the database as well as the methodology used in the construction of RPPIs. In the same section, we will seek, briefly, to draw attention to the great challenges in constructing these indices, comparing the Portuguese situation with some practices in some European Union countries. Section 3 is devoted to the presentation of results and discussion. Beyond the hedonic model in its three versions, and the respective spatial data analysis, the results will also be presented with reference to the central price tendency (mean and median price). Although this last approach may lead to biased results, it remains very attractive for its simplicity. Section 4 is dedicated to the spatial data analysis. The findings and conclusion will be presented in Section 5, in which a set of recommendations will be made, seeking to contribute to the elaboration of a strategy for regular publication of qualified RPPIs by public authorities.

2. Methodology and data

Something has already been said about the importance of good price indexes for the housing market. Good RPPIs has numerous applications in various fields. It is a good indicator of macro-economic activity in a country or a region. It can also serve as an indicator for the decision-maker and the responsible of monitoring the monetary policy. Good RPPIs are necessary to evaluate the wealth of families for whom, as we have seen, the house is in many cases the most valuable active. It can also make a valuable contribution in the measurement of inflation as key input for a basket of consumer prices. Finally it gives security and increases the efficiency of the housing market by introducing credible information, independent and timely.

Despite this importance, the property market in Portugal today still lacks a system of pricing information that gives agents transparent conditions and meets the other functions listed above. There is no systematic collection of transaction price by the national statistical system public. The INE (National Statistics Institute) publishes the monthly "Survey on bank evaluation on housing" trying to build leading indicators of housing prices. The values collected in these surveys, despite not being formal prices, represents the intention of purchase, thus constituting a reasonable approximation. The information is provided by a set of credit institutions operating in Portugal, considered as the most representative of the market for granting housing loans. The information is transmitted on a monthly basis and is used to calculate basic indicators of the average property assessment for a particular typology, $\bar{Y}_{ti}$ in period t, and in the territorial unit i, using the following formula:

$$\bar{Y}_{ti} = \sqrt[n]{\prod_{j=1}^{n} {}^{Y_{tij}}\!/\!_{A_{tij}}}, \qquad (1)$$

Being Y_{tij} the value of the assessment of the property j with the given typology j in period t in the territorial unit i; A_{tij} the corresponding net area of the property j in period t, in the territorial unit i, and finally, n, the number of housing assessments.

Beyond this indicator INE, in Portugal there are two other sources of statistics on housing prices produced by private entities. The first, ImoEconometrics, develops a set of 103 housing price indexes, thus intending to follow the evolution of the various housing markets in continental Portugal. The information that feed the indices constructed is collected from the INE Survey on bank evaluation on housing. The “Confidencial Imobiliário” is the second source of housing prices. Its price index Ci (Confidential Estate Index) aggregates information from real estate agencies with nationwide coverage. Beyond Confidential Estate Index, it also publishes indexes micro-scale at the parish level and also segmented by dwelling type (houses and apartments). The access to data and all the respective methodology is not free.

The construction of price indices for housing (RPPIs) raises several challenges, as we said, due to the particularities of the real estate market. The first has to do with the data source. The second has to do with methodological issues and seeks to answer important question as the need to stratify geographic unit or to identify sub-markets of similar dwelling, or still, the choice of the most appropriate econometric techniques, the variables to select or the most appropriate functional form. A regular publication of RPPIS implies the existence of a solid and reliable organization capable of promoting the collection of information, guaranteeing their quality and building the indexes properly. There are several types of database. Their distinctions depend on the time of collection which can be done at various stages of the process of buying and selling a property. We can use the prices advertised in newspapers or specialized agencies. We can use at a later stage of the transaction the evaluation of the credit institution that lends the money to the buyer. We can finally use transaction prices from the purchase documentation. These prices are normally collected from the notary authorities who are responsible for the legal properties registration. We mention here only two examples. The first example concerns the index "Notaires-INSEE," published in France since 2002, from an agreement between the INSEE ("Institut National de Statistique et des Etudes Economique") and the superior council of notaries ("Conseil Supérieur du Notariat ").[1] During the sales registration, all the descriptive information of the property is loaded into a database (BIEN, Base d'Informations Economique Notary) feeding later the construction of the price index. The second example takes us to the UK where the Halifax index is published since 1983 (Fleming & Nellis, 1984). Being initially built from simple weighted average, this price index housing has quickly evolved into hedonic methods estimated from a database built by the bank that gives the name to the index upon the requests for credit for housing purchase. Note that the Halifax index includes cases in which the sale is not realized.

2.1. Description of the data

In our present exercise, our database covers a geographical area comprising two counties, Aveiro and Ílhavo, in which there are approximately 110 000 inhabitants in an area of about 275 km2. Our database was built from the National Real Estate Portal - Casa Sapo. We collected and further processed 14087 observations covering a period between 2003 and 2010. In this treatment, we had to eliminated several hundred observations, one omitting relevant data that prevented their treatment, and other (outliers) presenting values far away from the average distribution. The database contains pricing announced (which may or may not correspond to a sale). The database covers villas (house) as well as apartments, new and used and in different locations. It also has a set of variables corresponding to attributes or characteristics of the house. There is an important limitation that should be noted and that has to do with the paucity of data for the earlier years of the sample and particularly with the 2004 data. This shortage of observation reveals an important methodological issue revealing the importance of hedonic method and in particular the time dummy variable methods that allow precise estimations provide there is good quality data.

[1] For a formal methodology description see (Dubujet, 2000)

Table 1: database description

Year	2003	2004	2005	2006	2007	2008	2009	2010
Houses	45	1	41	168	875	798	799	932
Apartments	164	17	189	571	2606	1899	2583	2399

The construction of an index of housing prices from a particular database must go through several steps. The first involves the breakdown of typological and geographical units, setting different zones and sub-markets. The second step consists of introducing corrective elements for the qualitative effects of each observation, and estimate prices in each sub-market. The price indexes are constructed in a third step according to previously established weighting criteria (Laspeyres, Paasche, Fisher and others).

2.2. Stratification method

Note that the stratification method bypasses the second stage, i.e., it does not require any econometric estimation. In this sense, it represents the less data intensive method. It just needs information about the price, type and geographic location. Despite the potential effects of estimation bias arising from the non-randomness of the sample and the evolution of the quality of the stock of real estate, we start applying this methodology in a first approach because of its validity arising mainly from its simplicity. For the stratification process of our sample we also seeks to apply simple and intelligible criteria that can be replicated in any other geographical area of the country and simultaneously guarantee a certain homogeneity in the evolution of prices. Regarding the geographical stratification we will use the new version of the Typology of Urban Area (TIPAU) published in 2009 by the National Institute of Statistics applied to parishes. Like the previous version, the 2009 TIPAU is a tripartite classification of the parishes of the Portuguese territory in predominantly urban areas (APU), medium urban areas (AMU) and predominantly rural areas (APR):

Predominantly urban area (APU): parish that includes at least one of the following requirements: 1) the higher value between the average of the weight of the urban resident population in the total population of the parish and the weight of the urban area in the total area of the parish exceed 50% 2) the parish includes the headquarters of the City Council and has a resident population of more than 5,000 inhabitants; 3) all or part of the parish partially includes a place with a resident population of more than 5,000 inhabitants, and the weight of the population of the place in the total population resident in the parish or in the total population of the place is more than 50%.

Moderately urban area (AMU): parish that includes at least one of the following requirements: 1) the higher of the average of the weight of the resident population in the total population of the parish and the weight of the area in the total area of the parish corresponds to the Urban Space , and the weight of the area is predominantly rural space occupancy exceeds 50% of the total area of the parish, and 2) the higher of the average of the weight of the resident population in the total population of the parish and the weight of the area in the total area of the parish corresponds the urban space in conjunction with semi-urban space, and the weight of the space area predominantly rural occupation shall not exceed 50% of the total area of the parish, 3) integrates the parish seat of the city Council and has a resident population equal or less than 5,000 inhabitants, 4) integrates parish wholly or partially a place with a resident population more than 2,000 inhabitants and less than 5,000 inhabitants, and the weight of the population of the place in the total population resident in the parish or in the total population residing in place is equal to or greater than 50%.

Predominantly rural area (APR): parish is not classified as "Predominantly Urban Area" or "moderately Urban Area".

The Table 2 shows the parishes of both municipalities indicating their typology.

Concelho	Freguesia	Tipo
Aveiro	Aradas	APU
Aveiro	Cacia	AMU
Aveiro	Eirol	AMU
Aveiro	Eixo	AMU
Aveiro	Esgueira	APU
Aveiro	Glória	APU
Aveiro	Nariz	AMU
Aveiro	Nossa Senhora de Fátima	AMU
Aveiro	Oliveirinha	AMU
Aveiro	Requeixo	APR
Aveiro	Santa Joana	APU
Aveiro	São Bernardo	APU
Aveiro	São Jacinto	APR
Aveiro	Vera Cruz	APU
Ílhavo	Gafanha do Carmo	APR
Ílhavo	Gafanha da Encarnação	APU
Ílhavo	Gafanha da Nazaré	APU
Ílhavo	Ílhavo (São Salvador)	APU

Table 2: Parishes classification

For the typological stratification we disaggregate the sample into six distinct types: villas, apartments TO, T1, T2, T3 and T4. The shortage of housing observations prevented us from considering other types of housing. Finally, we set three price indexes according to various criteria aggregation: an index for villas, apartments, and a global index. As for the weighing criteria, we successively use the Laspeyres, Paasche and Fisher methods. In the first case, we tie up the implicit quantities of the base-year and observe the evolution of the price of the resulting basket. In the case of the Paasche index, we use the implicit quantities of the current year and compare it using current price against the prices of the base year. As for the Fisher index, it is simply the square root of the product of the Laspeyres and Paasche indexes.

Laspeyres Price Index = $$\frac{\sum_{i=1}^{n} P_i^{t+1} Q_i^t}{\sum_{i=1}^{n} P_i^t Q_i^t} \qquad (2)$$

Paasche Price Index = $$\frac{\sum_{i=1}^{n} P_i^{t+1} Q_i^{t+1}}{\sum_{i=1}^{n} P_i^t Q_i^{t+1}} \qquad (3)$$

Fisher Price Index = $$\sqrt{\frac{\sum_{i=1}^{n} P_i^{t+1} Q_i^t}{\sum_{i=1}^{n} P_i^t Q_i^t} \times \frac{\sum_{i=1}^{n} P_i^{t+1} Q_i^{t+1}}{\sum_{i=1}^{n} P_i^t Q_i^{t+1}}} \times \sqrt{\frac{\sum_{i=1}^{n} P_i^t Q_i^{t+1}}{\sum_{i=1}^{n} P_i^t Q_i^t} \times \frac{\sum_{i=1}^{n} P_i^{t+1} Q_i^{t+1}}{\sum_{i=1}^{n} P_i^{t+1} Q_i^t}} \qquad (4)$$

2.3. Econometric Approach

In addition to the stratification method using the mean or median prices, we mentioned two other more sophisticated methods: hedonic methods and repeat sales. The starting point of the hedonic method is simple and based on the assumption according to which certain goods are demanded in the market not by the good itself but rather by the intensity of the characteristics that define it. Thus, the search of a home is determined not by the house itself but by its attributes which are many and various: location, rooms, equipment, garage etc. (Maleyre, 1997). The first hedonic regression is usually attributed to (Waugh, 1928) who had the idea of regressing the price of a bundle of asparagus based on three characteristics: the color, the rode and tip size. Later works (Griliches, 1971; Lancaster, 1966; Rosen, 1974) generalized this approach to the social sciences and in particular to the economy. In this paper, we use the hedonic method, which assumes previously that the price of a house depends on a combination of a limited set of qualitative and quantitative attributes. The relationship between the price per m2 of housing and their respective attributes is estimated from a regression based on a set of observations of actual prices, and then used to reconstruct the price of a dwelling in the same stratum reference sample. Thus the choice of the relevant explanatory variables to include in the model is crucial. We will consider successively some qualitative and quantitative variables that characterize the housing variables and the respective location and finally some temporal variable. Again we highlight here the parsimony criterion that determines the necessary compromise between the explanatory power of the model and its simplicity and applicability to local, regional and national levels. It is also important to compare our results, built according to the available data with the experience of other countries, which lead us to suggest the inclusion of other relevant variables to include in future developments.

Concerning the qualitative and quantitative variables, the Halifax House Price Indices are derived from information on the following house characteristics (Fleming & Nellis, 1984):

- Purchase price.
- Location (region).
- Type of property: house, sub-classified according to whether detached, semi-detached or terraced, bungalow, flat.
- Age of the property.
- Tenure: freehold, leasehold, feudal.
- Number of rooms: habitable rooms, bedrooms, living-rooms, bathrooms.
- Number of separate toilets.
- Central heating: none, full, partial.
- Number of garages and garage spaces.
- Garden.
- Land area if greater than one acre.
- Road charge liability.

The Notaires-INSEE indices published in France, in the respective equations, use the following descriptive variables of the property (Dubujet, 2000):

- Number of bathrooms

- Number of garages
- Floors
- Lift
- Number of rooms
- Surface by room
- Net Surface
- Land area (for housing)

In the present study and considering a multiple set of possible variables, we introduced in our models the following characteristics:

- Net Surface
- Number of bedrooms
- Presence of storage room
- Garage
- Land area (for housing)
- Garden (for housing)
- Detached (for housing)
- Parish typology
- New / Used

Many other variables were excluded because they were not statistically significant, while others were not included due to lack of availability. The floor or the presence of lift appear to be two good examples and would certainly help to improve the model as suggested by the two examples referred.

Regarding the variables of location and time, there are numerous solutions that often are subordinated to the existing database and the type of information provided. In the present study, we chose, as already mentioned, the new version of the Typology of Urban Area (TIPAU) published in 2009 by the National Statistics Institute. As such, we introduce a dummy, APU, including the house located in areas predominantly of urban typo, joining together the other two remaining typologies (AMU e APR), i.e., the parishes with moderately urban and predominantly rural areas. Many other models have more complex location variables related with accessibility or proximity to central areas or equipment. These are certainly relevant for more precise estimation but we consider that its application is not feasible for the construction of a national RPPI since its replication would put numerous practical problems. For temporal variables we included time dummies for each year covering the period under review.

Regarding the econometric model, the literature offers different functional forms.[2] The linear model is the simplest functional form using the values in level in both the price and the explanatory variables. The log-linear model uses the logarithm of the variables on both sides of the equation while the semi-log model uses only the price logarithm maintaining the explanatory variables in level. The Box-Cox transformation (Box & Cox, 1964) is generally regarded as a flexible functional well suited to estimate hedonic models. It represents a family of exponential transformations that generalizes traditional options such as those just mentioned. Originally, the Box-Cox transformation aimed to simultaneous resolute three problems related to the violation of fundamental assumptions of OLS model: the linearity of the parameters, the normality of variables and homoscedasticity. The transformation introduces a new parameter ʎ according to the following expression:

$$z = y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & if\ \ \lambda \neq 0 \\ \ln(y) & if\ \ \lambda = 0 \end{cases} \qquad (5)$$

Although the transformation does not guarantee the normality of the transformed data, it reduces several problems related to the estimation, prediction and inference (Hyde, 1999). As it is easy to see, this transformation incorporates many cases traditionally used in the literature (Osborne, 2010):

- ʎ=1.00: original data (in this case there is no transformation)
- ʎ=0.00: logarithmic transformation

---

[2] For a good review of the literature on the subject see (Cropper, Deck, & McConnell, 1988)

- ʌ=0.50: square-root transformation etc. ..

(Taylor, 2003) claims that the linear model is not generally adapted to house pricing estimation since the effect of a given attribute is not linear, i.e., it is not constant, varying with the level of the same attribute. In this sense we introduce the Box-Cox transformation in our model applying it in the dependent variable. The estimation of the parameter ʌ will be made using the method of maximum likelihood.[3] In this case, the impact of a unit change in a particular variable with a shadow price β on the price of a dwelling with a starting price p_0, measured in percentage change is given by the expression:

$$g = 100\left[\left(1+\frac{\lambda\beta}{p_0}\right)^{\frac{1}{\lambda}}-1\right] \qquad (6)$$

Thus, our model, in the Time Dummy Variable Method version, will be based on a hedonic regression including year dummies, where the dependent variable is subjected to a box-cox transformation. The estimation focuses on all the n observations of all periods. The equation takes the following form:

$$p_n(\lambda) = \beta_0 + \sum_{\tau=1}^{T} \delta^{\tau} D_n^{\tau} + \sum_{k=1}^{K} \beta_k x_{nk} + u_n \qquad (7)$$

In which, D_n^{τ} represents the time dummies variables and x_{nk} the k independent variables of the model described above.

The econometric approach is associated with a set of recurring problems. The first has to do with the choice of variables or more particularly with the possibility of missing important variables leading to biased parameter estimations. In the specific case of RPPIs where the need for parsimony models severely limits the number and variety of variables to be included, it is reasonable to assume the inevitability of this kind of bias, being difficult to measure the sign and magnitude of this effect (Eurostat, 2011). Spatial dependence and its treatment represent another line of investigation that will be specified in a later section of this paper. Finally there is the multicollinearity, a recurring problem in hedonic regressions, due to the correlation between the explanatory

[3] For the estimation we follow (Hyde, 1999).

variables. In the present case, in which our main focus is the production price index and not the estimation of shadow prices, the multicollinearity and the consequent instability in the estimated values of the parameters do not represent a major concern since we are mainly focus on price rather than shadow prices (Eurostat, 2011).

3. Results

In the introduction, we referred four methodologies used in the construction of price indexes in real estate: central tendency method (stratification method), hedonic method, repeated price and appraisal methods. In this chapter we will deal with three of these methodologies. We will use the INE (National Statistics Institute) "Survey on bank evaluation on housing" as a benchmark to compare our results with those concerning the INE about the municipality of Aveiro. Firstly, we calculate the price indexes for our database area comprising the Aveiro and Ílhavo municipalities, using the stratification methods. Secondly, we proceed with the estimations carried out from different hedonic models.

The average values of inquiries to bank assessment on housing published by the INE are measured in euros per m2. The calculations are made for different typologies and for various levels of geographical disaggregation. Prices exist for all NUTSIII (regional level), for the two major metropolitan areas of Lisbon and Porto and, finally, also to urban areas and counties that include medium-sized cities. Thus, we used from the INE database the average values of bank evaluation on housing comprising the municipality of Aveiro in a quarterly series covering the period from the third quarter 2003 and 4th quarter 2009. The average values correspond to the two sub-market apartments and townhouses, and their aggregation are made according to the method explained above. The area does not correspond exactly to our sample since it does not include the county of Ílhavo.

We can foresee in Figure 1 a period of high prices between 2003 and 2005, with a peak in the third quarter of 2004. From there, we have a first decrease in the first quarter of 2006 followed by a period of stability until 2007. The years 2008 and 2009 are marked by a downward trend in prices with a slight fall in the second quarter of 2009. This pattern is followed in a more or less similar way for both sub-markets (villas and apartments). We will explore among the following two sections, first the methodology for stratification and then the hedonic estimations.

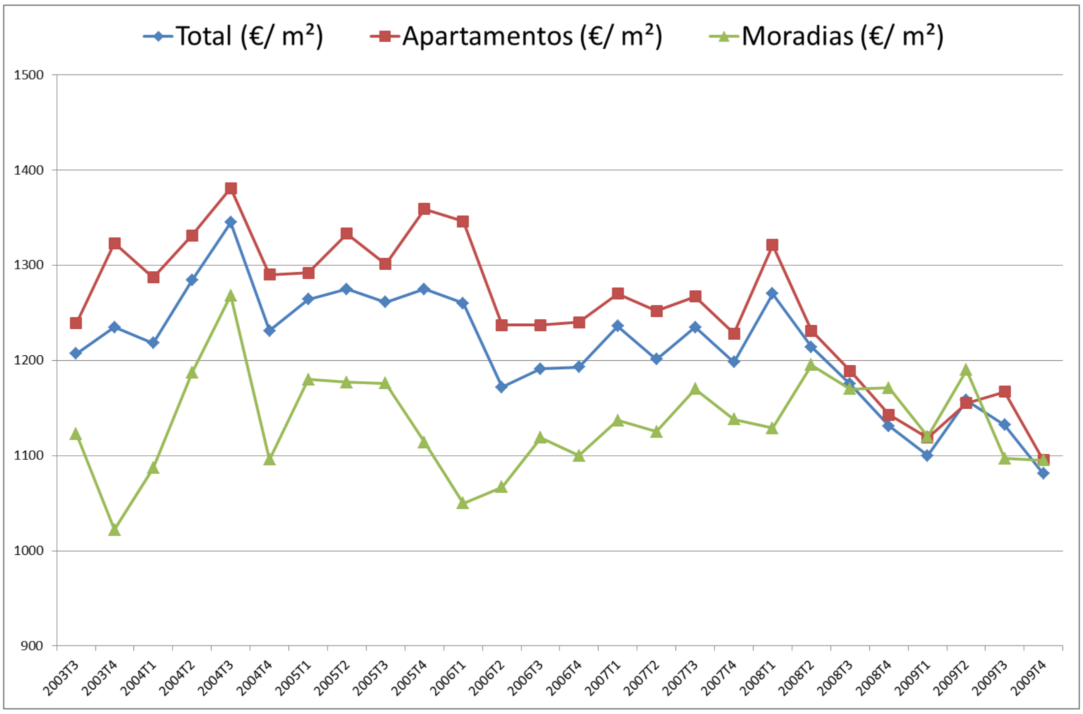

Figure 1: Appraisal method

3.1. Stratification Methods

These methods, as we have already noted are those that are less intensive data. Apart from the price and area, only the location is needed.[4] According to the degree and type of stratification used, we may also need the house typology.

[4] The net area corresponds to the surface of the house (including lobbies, corridors, bathrooms, and other storage compartments and cupboards) measured by the inner perimeter of the walls, discounting all thralls up to 30cm, interior walls, partitions and ducts.

In this exercise, we are going to calculate average prices separately for villas and apartments, and disaggregate the flat in T0, T1, T3 and T4. We distinguish two geographic areas, the first comprising the parishes with predominantly urban areas (APU) and the second with the parishes medium urban areas and predominantly rural (AMU e APR). We present for each sub-market their respective mean and median. However, for the aggregation of prices and respective RPPIs construction, we will use the median, since the majority of price distributions exhibit positive asymmetries.[5] A Figure 2 illustrates the price distribution of apartments sold in Q3 2009, thus confirming, as an example, the existence of a positive asymmetry coefficient. In those situations where the observations tend to accumulate in the left distribution and as it is recommended in the literature, the median becomes more representative of the population, mitigating the effects of outliers.

We build separate indexes for the apartments and the villas, both adding to the overall index.

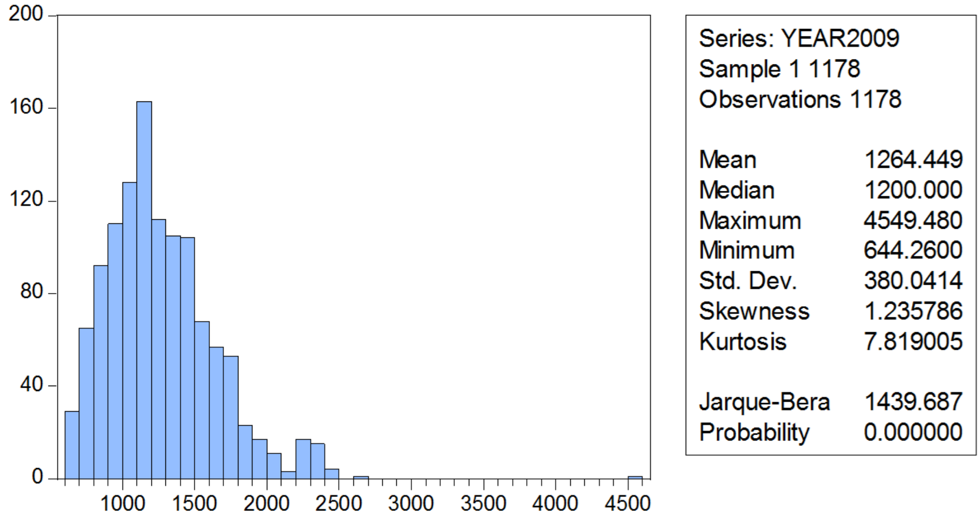

Figure 2: distribution of apartments' prices in 2009

[5] Since house price distributions are generally positively skewed (predominantly reflecting the heterogeneous nature of housing, the positive skew in income distributions and the zero lower bound on transaction prices), the median is typically used rather than the mean (Prasad & Richards, 2006).

3.1.1. Apartments

Table 3 and Table 4 presents a set of descriptive statistics relative to the prices of apartments in 2003 and 2010 respectively in predominantly urban areas (APU, Table 3) and the remaining (AMU and APR, Table 4). One of the advantages of the stratification method is its ability to disaggregate information according to the type and geographic area. However, we need a certain quantities of observation in order to reach a minimum level of representativeness. As shown in the tables, the number of observations is reduced especially in the early years of the study period and also outside the predominantly urban areas. In this sense, the calculations for certain strata stands as a theoretical exercise, and as such, may lack of statistical validity, namely for the price of the sub-markets where information is scarce.

Comparing the mean and median, we find that, as indicated by the coefficients of asymmetry, the second surpasses the first in most cases. Such is the cases of T0 in predominantly urban areas from 2008 onward, whose price distribution has a negative skewness and where the median surpasses the average in 2009 and 2010. We note also that, concerning the typology and years with more than 100 observations, the standard deviation keeps relatively constant over time with values ranging between 300 and 400 euros.

AMU + APR								
	2003	2004	2005	2006	2007	2008	2009	2010
	T0							
Observations	2	0	0	0	0	2	3	3
Standard deviation	219.31					579.43	639.46	584.27
Skewness							1.73	1.73
Mean	1111.10					1465.28	1232.21	1200.34
Per cent change							-15.91%	-2.59%
Median	1111.10					1465.28	863.01	863.01
Per cent change							-41.10%	0.00%
	T1							
Observations	1	1	8	18	30	32	41	45
Standard deviation			51.98	321.60	427.92	522.52	425.48	271.81
Skewness			0.57	1.92	2.02	2.58	3.03	1.80
Mean	859.04	927.24	1020.19	1325.57	1317.48	1264.82	1302.97	1138.37
Per cent change		7.94%	10.02%	29.93%	-0.61%	-4.00%	3.02%	-12.63%
Median	859.04	927.24	1000.00	1238.57	1162.28	1133.33	1200.00	1141.43
Per cent change		7.94%	7.85%	23.86%	-6.16%	-2.49%	5.88%	-4.88%
	T2							
Observations	17	6	12	29	114	85	71	77
Standard deviation	190.82	448.98	109.02	169.30	184.90	504.01	300.27	381.48
Skewness	0.61	-0.67	1.88	2.59	2.34	3.76	3.08	3.13
Mean	902.57	1441.19	972.45	961.54	1017.63	1089.66	930.95	995.92
Per cent change		59.68%	-32.52%	-1.12%	5.83%	7.08%	-14.56%	6.98%
Median	897.84	1625.00	958.82	954.55	1000.00	952.38	842.11	950.00
Per cent change		80.99%	-41.00%	-0.45%	4.76%	-4.76%	-11.58%	12.81%
	T3							
Observations	6	1	10	18	84	63	78	68
Standard deviation	87.28		89.59	116.51	150.16	379.05	149.12	183.15
Skewness	0.60		-0.15	0.40	0.09	3.56	1.91	1.43
Mean	785.97	848.00	920.70	930.39	872.79	966.80	856.90	822.74
Per cent change		7.89%	8.57%	1.05%	-6.19%	10.77%	-11.37%	-3.99%
Median	761.13	848.00	918.33	922.95	875.00	905.98	813.66	797.73
Per cent change		11.41%	8.29%	0.50%	-5.20%	3.54%	-10.19%	-1.96%
	T4							
Observations	5	0	1	1	5	5	1	0
Standard deviation	256.99				110.10	154.48		
Skewness	1.04				0.53	-0.27		
Mean	893.79		964.29	966.67	844.66	635.36	482.86	
Per cent change				0.25%	-12.62%	-24.78%	-24.00%	
Median	863.30		964.29	966.67	784.41	687.50	482.86	
Per cent change				0.25%	-18.85%	-12.35%	-29.77%	

Table 4: Descriptive statistics in AMU+APR areas

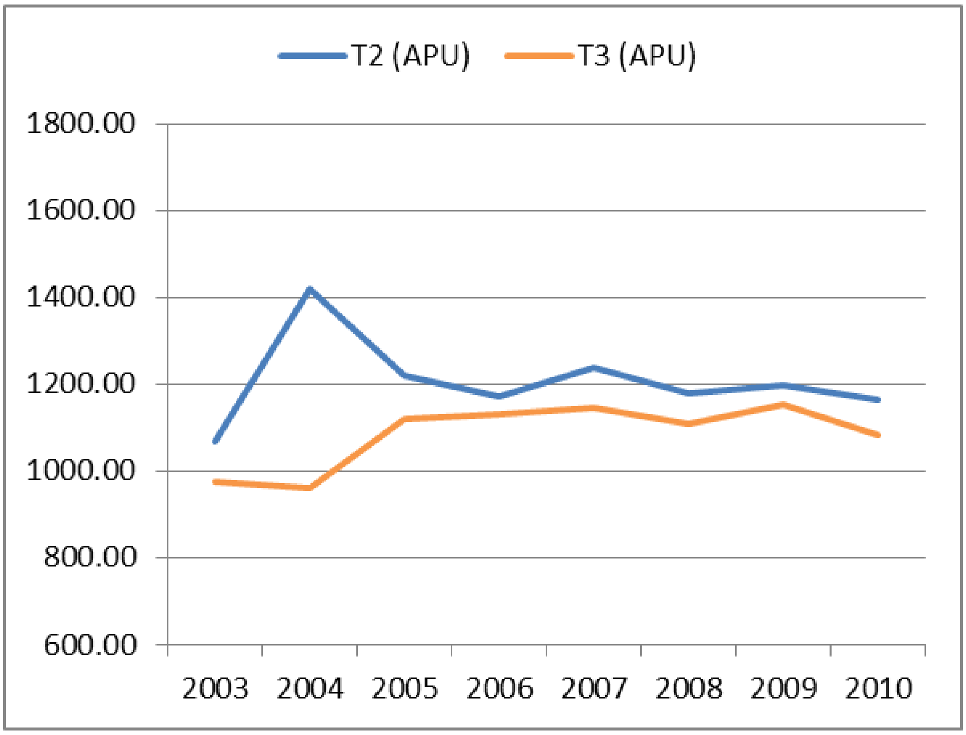

Figure 3: evolution of median price of T2 and T3 apartments (urban areas)

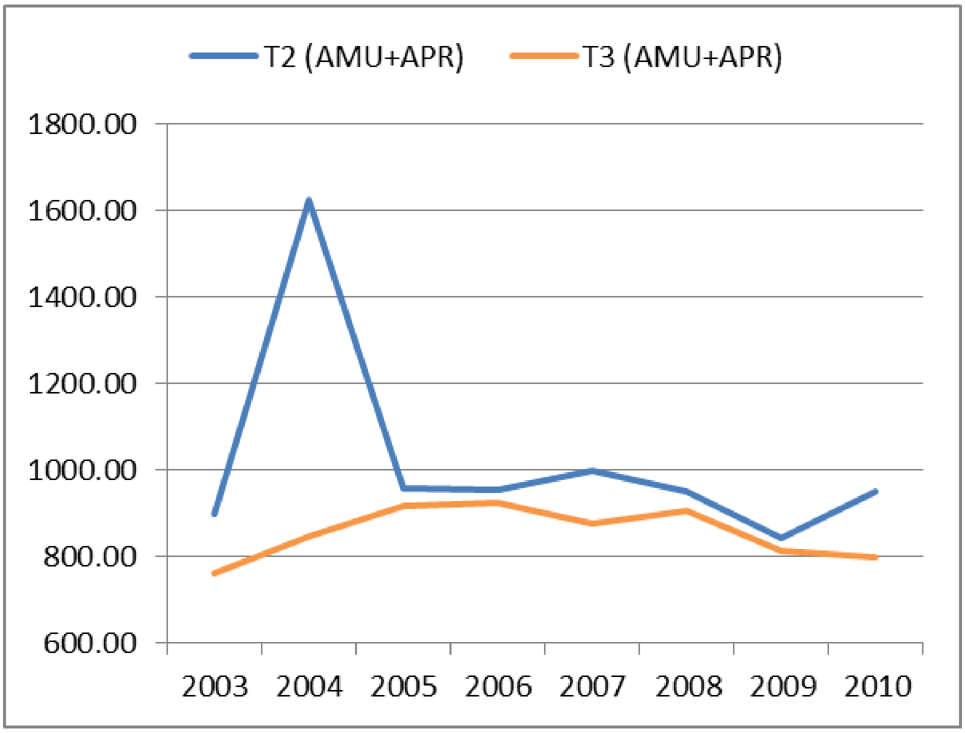

Figure 4: evolution of median price of T2 and T3 apartments (semi-urban and rural areas)

Comparing the growth rates applied to the mean or the median, we observe some significant differences, tending however to vanish as the number of observations increases. In some periods and in certain typologies, the growth rate of the mean exceeds the median while the opposite happening in other groups. However, both tend to be consistently of the same sign.

Figure 3 shows the evolution of the median price of T2 and T3 in predominantly urban areas between 2003 and 2010, thus representing an example of the degree of disaggregation of information that can be displayed with the stratification method. Provide there are enough degrees of freedom, the typological and geographical segregation allows a high degree of homogeneity in the stratum studied. In the particular case of Figure 3 we observe an upward trend of prices during the early part of the period, with a peak in 2004, in the case of T2, followed by a period of stability between 2006 and 2009, a slight fall onward.

Figure 4 shows the same curve for the same types, but corresponding to medium urban areas and predominantly rural. The pattern is similar, with an initial rise, followed by stabilization and ending with a downward trend of prices. The peak of 2004 on apartments T2 is much more pronounced. On the other hand, we have, in 2010, a slight increase revealing a countercyclical trend relative to the other typologies and geographical units. We also noted that the downward trend starts earlier compared with the predominantly urban areas.

As a last record, we may point to the fact that 2010 prices remain above the 2003 prices. Although this still corresponds to a drop in real terms, it demonstrates the slow adjustment of real estate prices in view of the outbreak in 2007 of the current financial crisis with its epicenter in the property sector.

We proceed with the construction of a RPPI that aggregates all the apartments over the period 2003-2010. Once defined all strata based on geographic and typological criteria, we obtain the respective medians (as the best representative measure of each stratum), and finally we aggregate the prices according to the weighting criteria described above.

To calculate the price indexes, be that the Laspeyres, the Paasche or Fisher method, we need prices and the implicit quantities. The latter represent the rule of proportionality between the volume and the prices and will be used as a reference to the basic elements weighting. Table 5 summarizes the information already provided above, with the median price for each stratum (area versus typology). Table 6 lists the quantities implied obtained by dividing the volume by the respective median.

	VOLUME				
	APU				
Year	T0	T1	T2	T3	T4
2003	1,323.34 €	22,451.94 €	72,168.02 €	44,546.70 €	3,171.57 €
2004	- €	2,952.33 €	6,979.40 €	2,797.31 €	- €
2005	11,137.14 €	42,108.51 €	99,106.38 €	42,751.64 €	1,833.33 €
2006	19,030.47 €	117,728.23 €	312,302.95 €	177,781.83 €	11,921.61 €
2007	70,903.70 €	642,516.09 €	1,423,398.73 €	780,848.28 €	129,481.96 €
2008	47,290.13 €	425,884.50 €	1,129,361.57 €	450,809.78 €	64,508.85 €
2009	56,414.54 €	612,496.26 €	1,488,353.87 €	726,126.76 €	147,855.88 €
2010	81,505.39 €	563,439.33 €	1,249,206.97 €	727,686.19 €	92,664.14 €
	AMU +APR				
Year	T0	T1	T2	T3	T4
2003	2,222.21 €	859.04 €	15,343.67 €	4,715.81 €	4,468.93 €
2004	- €	927.24 €	8,647.13 €	848.00 €	- €
2005	- €	8,161.50 €	11,669.41 €	9,207.04 €	964.29 €
2006	- €	23,860.21 €	27,884.57 €	16,747.08 €	966.67 €
2007	- €	39,524.33 €	116,010.36 €	73,314.29 €	4,223.30 €
2008	2,930.56 €	40,474.16 €	92,620.77 €	60,908.21 €	3,176.82 €
2009	1,970.59 €	53,421.57 €	66,097.75 €	66,837.89 €	482.86 €
2010	1,875.00 €	51,226.66 €	76,685.69 €	55,946.06 €	- €

Table 5: Apartment's prices using median price for each stratum (apartments).

	Implicit Quantities				
	APU				
Year	T0	T1	T2	T3	T4
2003	1.00	18.89	67.48	45.62	2.72
2004		2.00	4.91	2.91	
2005	7.21	32.95	81.25	38.09	1.00
2006	11.60	84.56	266.46	156.87	12.72
2007	44.78	449.94	1149.67	681.58	107.90
2008	30.67	329.91	956.80	405.80	62.55
2009	37.61	440.28	1240.29	629.46	125.58
2010	50.27	421.16	1070.75	671.71	83.84
	AMU +APR				
Year	T0	T1	T2	T3	T4
2003	2.00	1.00	17.09	6.20	5.18
2004		1.00	5.32	1.00	
2005		8.16	12.17	10.03	1.00
2006		19.26	29.21	18.15	1.00
2007		34.01	116.01	83.79	5.38
2008	2.00	35.71	97.25	67.23	4.62
2009	2.28	44.52	78.49	82.14	1.00
2010	2.17	44.88	80.72	70.13	

Table 6: Implicit quantities (apartments, stratification method).

The Laspeyres, the Paasche or the Fisher indexes are synthetic indexes in that they represent composite indicators of simpler index based on heterogeneous quantities that cannot be combined. When calculating weighted averages of these elementary indices, the choice of the weights determines the type of indexes referred earlier. The maintenance of the weights of the reference period corresponds to the Laspeyres index, while the Paasche index uses the weights of period t, comparing the price of this basket in period t with the price of the same basket in the base year.

The choice of one or the other has advantages and disadvantages. The Laspeyres index has two advantages (Balk, 1995). The first has to do with the fact that it is less demanding on statistical data, i.e. it only requires knowledge of the quantities of the initial period. Because the weights are constant the Laspeyres index is more stable against the Paasche index, since it varies solely as a function of price. The criticism that is usually made to the

Laspeyres has to do with the fact that it overstates the evolution of prices, to the extent that it does not incorporate the adaptation of the basket purchase by the agents. Indeed, the rising price of a particular good usually shifts demand toward any other substitute good. Thus, the basket of base-year tends to loose representativeness over time. In this sense, the Paasche index, being more demanding in the amount of data necessary for their calculation, continuously updated the price basket. However, the Paasche index, using the same argument but inversely, underestimate the inflation due to the biased weighting toward products whose price declined more compared to the average.

The property of circularity of indexes should ensure that they are not affected when comparing values more spaced in time (Goldfarb & Pardoux, 2011). That is, as the index series increases, the circularity prevents the occurrence of distortion, which implies that the basket of goods which are to be compared keeps its homogeneity. The chain indexes are the best example to illustrate the property of circularity. The chain indexes are calculated successively comparing a value with the one that immediately precedes it. In this sense they represent a chain of values, each one representing a specific period. The chain indices are not affected by distortionary effects over time (P. Hill, 1990).

The major criticism that is made to the Laspeyres Paasche and Fisher indexes derives precisely from the fact that they are not chain indexes. Used for short periods, they usually do not present problems of bias. However, for longer intervals, the composition of the basket, or more properly speaking, the real weighting elements changes and the distortionary effects can be significant. In this sense, the Laspeyres, Paasche and Fisher indices are often considered short term indexes. An often used solution involves the construction of series of chained Laspeyres indexes, published under the name chain Laspeyres indexes.

In our case, given the relatively short period under study, we will not worry for now, with this issue. However, in the section dedicated to hedonic methods, we will estimate yearly regressions in order to construct chain RPPIs.

Year	**Laspeyres Index**	**Paasche Index**	**Fisher Index**
2003	100.0000	100.0000	100.0000
2004	117.7230	137.2625	127.1179
2005	112.2469	113.2450	112.7448
2006	110.6734	113.0807	111.8706
2007	114.0653	116.4509	115.2519
2008	110.0993	110.3477	110.2234
2009	110.6057	113.5684	112.0773
2010	106.3151	110.1579	108.2195

Table 7: Apartment RPPIs, stratification methods

Table 7 shows the calculated indexes with different weighting and aggregation rules: Laspeyres, Paasche and Fisher. Figure 5 illustrates the evolution of each index over the period. All three evolve in parallel, portraying the same trend, with a rise in prices in 2004, followed by a period of stability between 2005 and 2009. The final part shows a common slight trend of falling prices. If we focus our attention to the Fisher index, we find that this downward trend actually begins a little earlier, in 2007. If we compare these curves to our benchmark (the INE "Survey on bank evaluation on housing" concerning the municipality of Aveiro for the same period, in Figure 1), we find similarly the peak of price in 2004, followed by the same period of stability, finalized by a trend-breaking prices starting from the fourth quarter 2007 and first quarter of 2008.

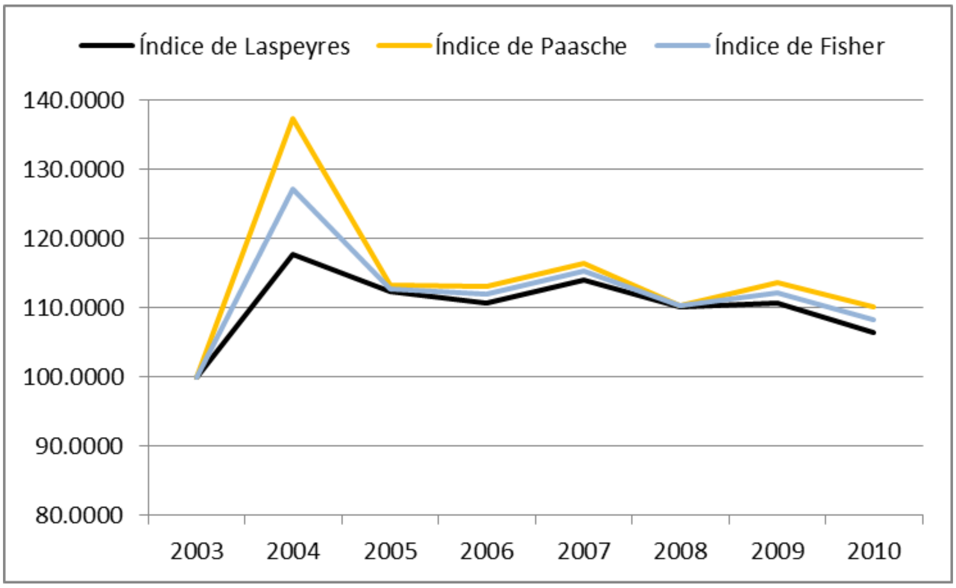

Figure 5: Apartment RPPIs, stratification methods

3.1.2. House (villas)

Regarding the house sub-market we reproduce the same methodology previously used for the apartments. However, given the small number of observations (3659) we cannot maintain the same level of disaggregation. We will not disaggregate by typologies, while maintaining the two geographical areas previously used, the first comprising the parishes with predominantly urban area (APU) and the second comprising the parishes averagely urban and predominantly rural (AMU and APR).

Figure 6 shows the pool distribution of price of all period. Once again we observe the positive asymmetry with prices tending to be concentrated at the left and lower of the price scale. Table 8 presents the same set of descriptive statistics. Again we signalize the lack of information that compromise the homogeneity of the sample and requires some caution in the interpretation of prices, with special emphasis in the early years of the study period. To calculate the price index, we calculate the implicit quantities based on the median prices (Table 9) and apply the Laspeyres, Paasche and Fisher weights methods.

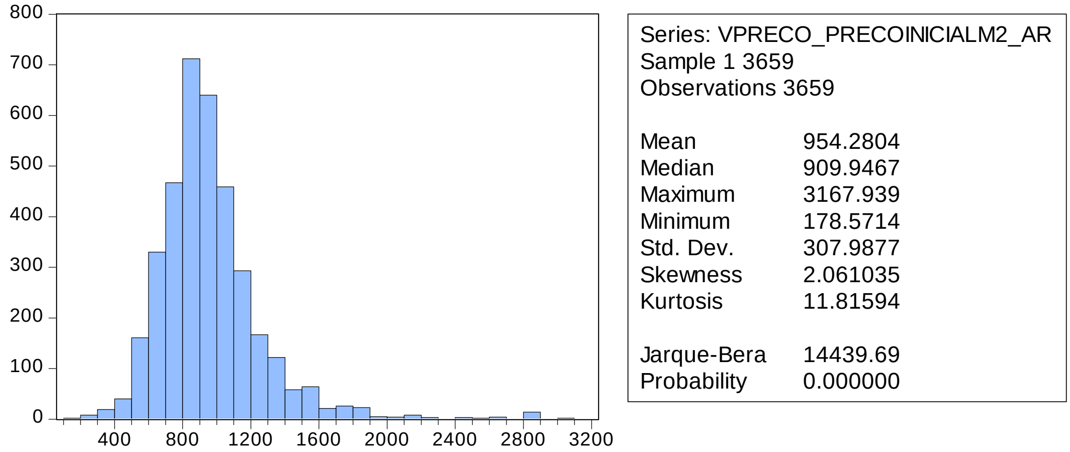

Figure 6: Distribution of House prices among all periods.

	APU							
Year	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Observations	28	1	29	126	651	640	569	707
Standard deviation	408.84	-	224.16	291.45	277.51	346.62	361.78	272.98
Skewness	3.88	-	1.00	1.36	1.72	2.22	2.70	1.39
Mean	836.67	833.33	865.00	952.13	954.43	1007.16	1006.78	993.12
Per cent change		-0.40%	3.80%	10.07%	0.24%	5.52%	-0.04%	-1.36%
Median	740.56	833.33	875.00	919.28	920.63	938.72	943.40	963.16
Per cent change		12.53%	5.00%	5.06%	0.15%	1.96%	0.50%	2.09%
	AMU + APR							
Observations	16	0	12	42	224	158	230	225
Standard deviation	287.10	-	191.22	175.85	284.58	264.79	265.21	257.24
Skewness	0.45	-	0.82	-0.35	1.87	1.83	1.23	1.20
Mean	784.86	-	763.90	832.56	841.76	840.08	887.85	880.55
Per cent change		-	-	8.99%	1.11%	-0.20%	5.69%	-0.82%
Median	772.82	-	760.00	836.31	815.34	826.10	855.26	838.93
Per cent change		-	-	10.04%	-2.51%	1.32%	3.53%	-1.91%

Table 8: Descriptive statistics in APU and AMU+APR areas (house sub-market).

	Volume							
Year	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
APU	23426.82	833.33	25085.01	119968.98	621333.85	644581.35	572856.27	702135.18
AMU + APR	12557.75	0.00	9166.80	34967.46	188554.74	132732.77	204204.85	198124.22
	Price (median)							
APU	740.56	833.33	875.00	919.28	920.63	938.72	943.40	963.16
AMU + APR	772.82	-	760.00	836.31	815.34	826.10	855.26	838.93
	Implicit quantities							
APU	31.63	1.00	28.67	130.50	674.90	686.66	607.23	728.99
AMU + APR	16.25	-	12.06	41.81	231.26	160.67	238.76	236.16

Table 9: Volume and implicit quantities for houses sub-market.

Year	Laspeyres Index	Paasche Index	Fisher Index
2003	100.0000	100.0000	100.0000
2004	112.5275	112.5275	112.5275
2005	111.2398	112.1091	111.6736
2006	118.5785	120.1449	119.3591
2007	117.7505	119.3606	118.5528
2008	119.8259	122.8593	121.3331
2009	121.5542	122.5245	122.0384
2010	122.5537	124.6250	123.5850

Table 10: House RPPIs, stratification methods

Table 10 contains the indexes calculated for the period 2003-2010. Figure 7 illustrates the respective evolution throughout all period. The three indices (Laspeyres, Paasche and Fisher) all evolve similarly, indicating, in the case of the municipalities of Aveiro and Ílhavo an upward trend throughout all period, differently from the apartments paths. If we compare this graph with the benchmark (the INE "Survey on bank evaluation on housing" concerning the municipality of Aveiro for the same period, in Figure 1), we can see that the curves relating to housing price index coincide with the INE survey for the period between 2003 and 2008, but does not follow the downward trend between 2008 and late 2009. Likewise, we not recognize in our results the 2004 peak and the instability between the third quarter of 2003 and first quarter of 2006. However, we again note that the number of observations in our sample is relatively scarce especially in the early years of the study period.

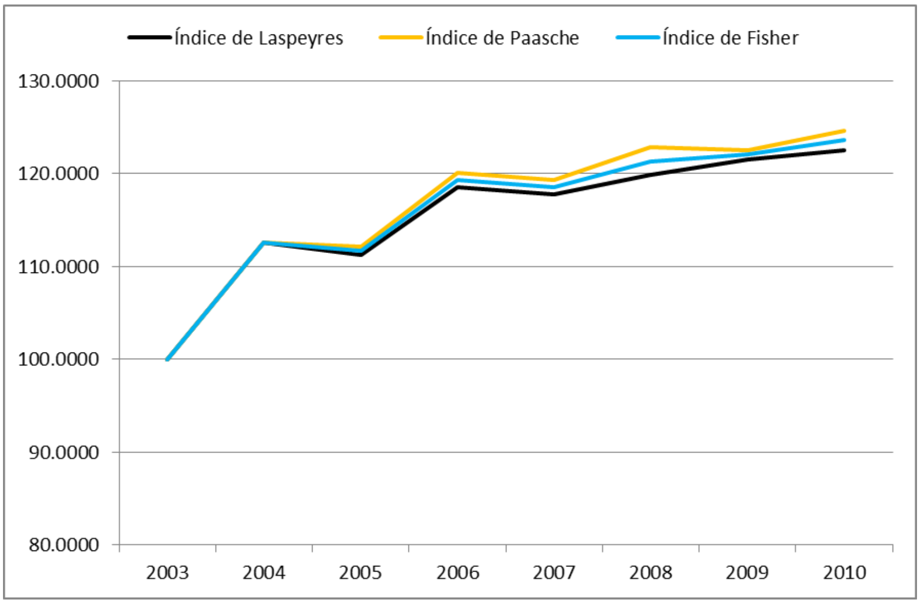

Figure 7: House RPPIs, stratification methods

3.1.3. Stratification methods: conclusion

We conclude this section with the construction of an index of real estate prices applied to all the strata including the two geographical levels (APU and AMU + APR) and 5 types (houses, apartments T1, T2, T3 and T4).

Year	Laspeyres Index	Paasche Index	Fisher Index
2003	100.0000	100.0000	100.0000
2004	117.5564	136.4694	126.6500
2005	112.1158	113.0972	112.6054
2006	112.0890	114.3458	113.2116
2007	114.7946	117.0268	115.9052
2008	112.5403	113.4877	113.0116
2009	112.7331	115.3086	114.0128
2010	110.1597	113.5831	111.8573

Table 11: RPPI global indexe

As it can be inferred from Table 1and Figure 8 the results for the global index is mainly influenced by the price of the apartments because of the respective representativeness in the sample. The path has already been described. As such it is not necessary to go further with the discussion.

Regarding specifically the method of indexes construction, this exercise reflects the potential of the stratification method as well as its limitations. Its simplicity is obvious, been easily understood by all agents. Moreover the level of disaggregation allows providing disaggregated information at the level of several sub-markets or segments of specific markets with relevant information for the market. However, this possibility is strongly conditioned by the availability of data. If the geographical and typological disaggregation allows working in theory with a good degree of sample homogeneity, the lack or scarcity of data may bias strongly the sample, compromising the representativeness of the housing stock and affecting the indexes quality.

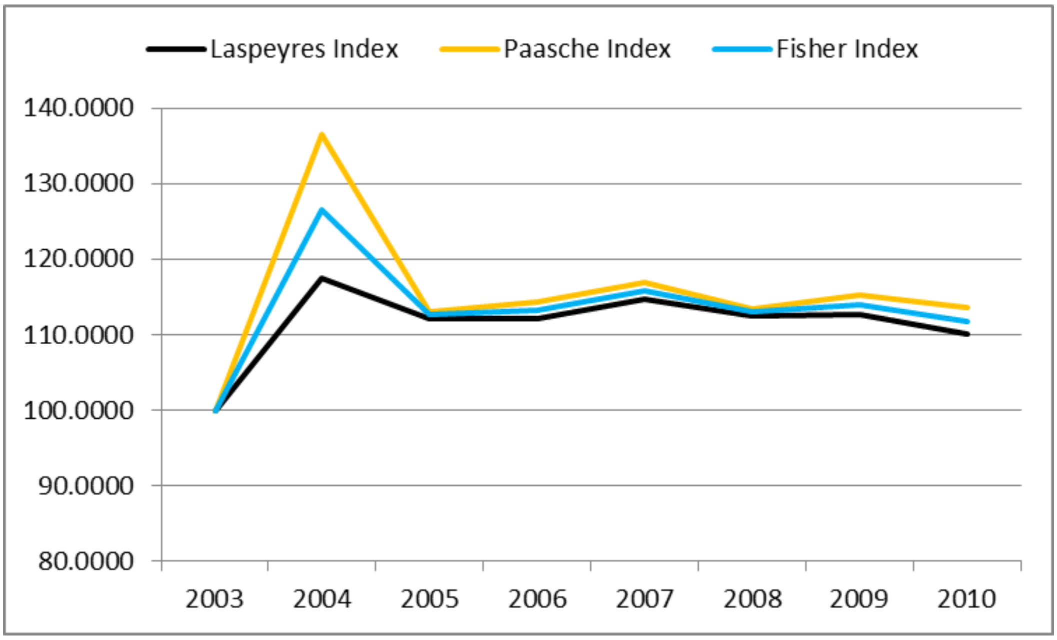

Figure 8: RPPI global index (graphic evolution)

3.2. Econometric approach

Among the econometric approach we will cover two variants of the hedonic approach: the time dummy variable method and the characteristic (or imputation) method. The hedonic approach, when compared to the methods previously studied, namely the stratification methods, is more data intensive, particularly in qualitative terms. Indeed, to calculate shadow prices, we need to know and quantify a set of characteristics of dwellings for which we intend to estimate their individual valuation. Nevertheless, in the particular case of real estate price indices and quoting the (Eurostat, 2011): "Although most hedonic regressions on house prices in the literature will often use many more explanatory variables, some studies show that reliable hedonic price indexes can be obtained with as few as four independent variables. "

3.2.1. The time dummy variable model

We begin by estimating the model expressed in equation (7), corresponding to a pooled regression over all period, thereby maximizing the degrees of freedom. As to the functional form, we followed (Hyde, 1999) and estimated the optimal ʎ parameter by a maximum likelihood function in order to approach our data and particularly the price per square meter of the normal distribution. We apply this methodology to the different submarkets, firstly with the apartment, and secondly with the houses. Finally we aggregate both in a global RPPI.

3.2.1.1 Apartments

The Figure 9, obtained from an algorithm described in (Hyde, 1999) and provided by the author in a Matlab code shows the graph of the log maximum likelihood, expressing also the limits of the 95% confidence interval constructed from the asymptotic distribution estimators.

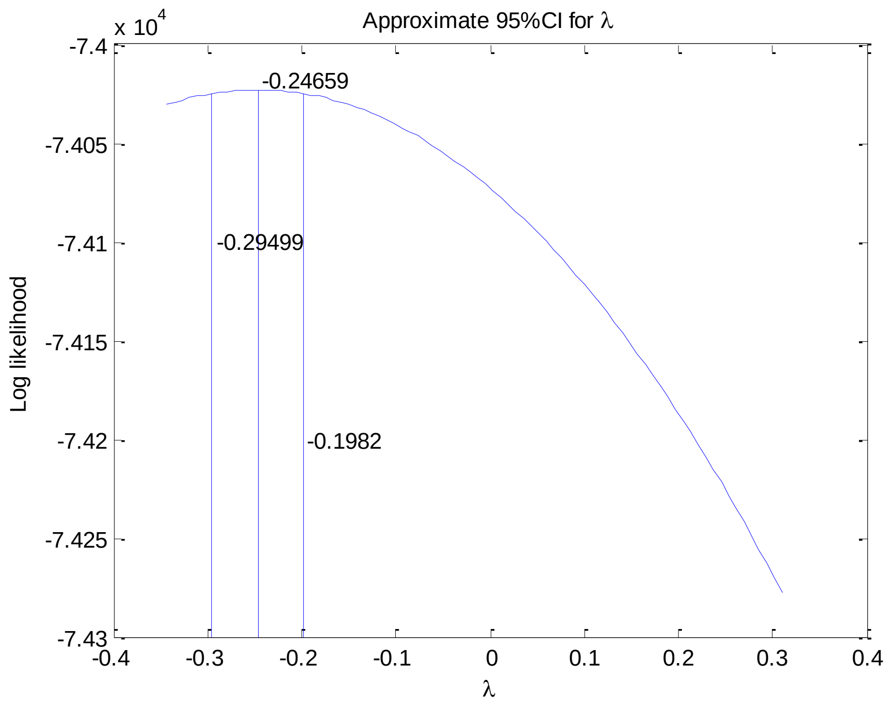

Figure 9: Matlab plot of the log likelihood function (Hyde, 1999) applied to the apartments.

According to the results, the optimum value of lambda is -0.24659. Note that the 95% confidence interval, between the range of -0.29499 and -0.1982 excluding the unit, thus justifying the transformation. Furthermore, it is also distinct from zero so the log transformation is not optimal.

With this transformation of the dependent variable, we proceed with the estimation of the time dummy variable model for the apartment's submarket. The results are shown in *Table 12*. As you can see, all explanatory variables are significant, including the 7 year dummies and the remaining six characteristics. The signal is consistent with the usual intuition, except for the number of storage ("VD04_ARRUMOS") that appears negative, although by a minor amount. The coefficient of determination is 31.6%.

Dependent Variable: PLAMBDA1
Method: Least Squares
Date: 05/21/13 Time: 11:22
Sample: 1 10428
Included observations: 10428
Newey-West HAC Standard Errors & Covariance (lag truncation=11)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.676023	0.016464	223.2726	0.0000
LOG(AREA)	-0.089957	0.003743	-24.03654	0.0000
QUARTOS	0.012820	0.001366	9.385458	0.0000
VD04_ARRUMOS	-0.006614	0.001419	-4.662425	0.0000
VD11_GARAGEM	0.006953	0.001679	4.141550	0.0000
VP02_NOVO	0.025696	0.001687	15.22859	0.0000
APU	0.038102	0.003220	11.83277	0.0000
DTA_2004	0.035555	0.013721	2.591354	0.0096
DTA_2005	0.016374	0.006587	2.485849	0.0129
DTA_2006	0.021631	0.006715	3.221093	0.0013
DTA_2007	0.021592	0.005761	3.747771	0.0002
DTA_2008	0.012915	0.005800	2.226822	0.0260
DTA_2009	0.015409	0.005895	2.613892	0.0090
DTA_2010	0.012123	0.005810	2.086563	0.0370

R-squared	0.316842	Mean dependent var	3.345234
Adjusted R-squared	0.315989	S.D. dependent var	0.051135
S.E. of regression	0.042291	Akaike info criterion	-3.487136
Sum squared resid	18.62587	Schwarz criterion	-3.477400
Log likelihood	18195.93	Hannan-Quinn criter.	-3.483847
F-statistic	371.5308	Durbin-Watson stat	1.069249
Prob(F-statistic)	0.000000		

***Table 12:* Pooled regression for all periods (apartments)**

The price indices are taken from the estimated parameters of the annual dummies, with the necessary adaptation designed to undo the Box-Cox transformation. By doing this we reach a series of price indices from 2003 (base-year) through 2010. The prices already transformed and normalized to 100 in 2003 are shown in Table 13.

Year	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
RPPI	100.00	122.48	109.33	113.23	113.10	107.64	109.44	107.32
% change	-	22.48%	-10.73%	3.57%	-0.12%	-4.83%	1.67%	-1.93%

Table 13: RPPI for apartments (time dummy method)

The Figure 10 illustrates the evolution of the RPPI for apartments between 2003 and 2010. We recognize a similar path when compared with the RPPIs calculated in the previous sections, with a major peak in 2004, with a rise of 22.8% over the previous year, followed by a period of relative stability and finally a downward trend in the prices starting in 2007 (-4.83% in 2008, 1.67% in 2009 and finally -1.93% in 2010).

.

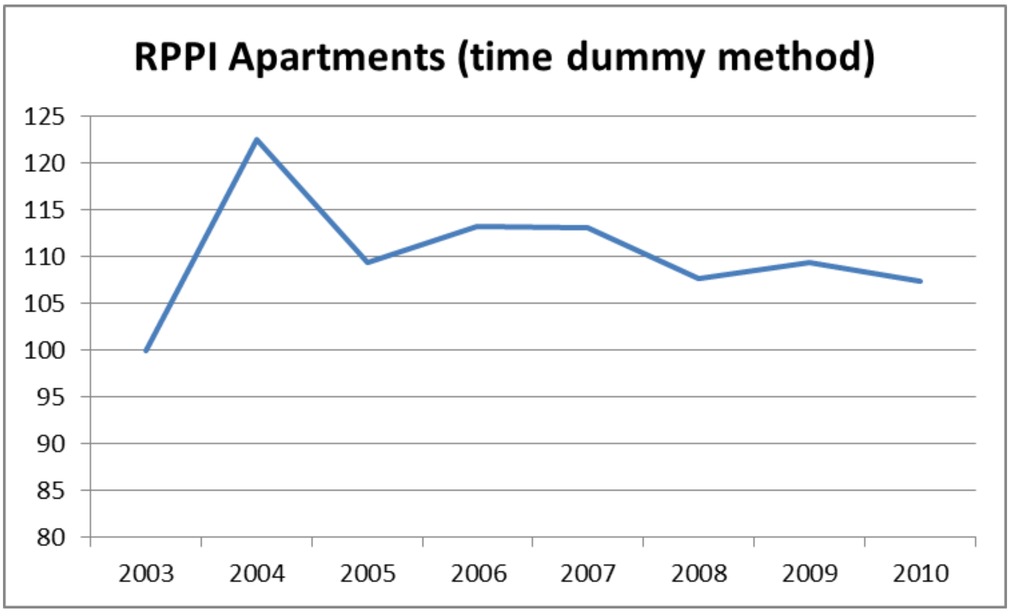

Figure 10: RPPI evolution (time dummy method): apartments

One of the problems related to this method frequently cited in the literature has to do with the fact that the series must be systematically revised as new periods are added. An additional year brings a new set of observations that affect the coefficients estimation and consequently the index of each year. On one hand, it may represent an advantage in that the increase of observations increases the efficiency of the estimators. But it is no less true that the systematic review of price indexes, even being small, would never be welcomed by real estate agents (Eurostat, 2011). One solution to this problem may be to estimate adjacent periods two by two with a single annual dummy, thus building a chain index, although decreasing significantly the degree of freedom.

3.2.1.2 Houses

Regarding housing, we tried to repeat the same methodology, estimating the optimal parameter for box-cox transformation and building the best model for the estimation of hedonic prices. As the dependent variable, we use the transformed price of per square meter. Concerning the explanatory variable we use the number of rooms, the presence and numbers of storage, the presence of garden, the fact of being detached or not, the fact that it is new or used, and finally the 7 annual dichotomous variables.

Dependent Variable: PLAMBDA2
Method: Least Squares
Date: 05/21/13 Time: 14:23
Sample: 1 3659
Included observations: 3659

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.542805	0.080452	106.1854	0.0000
LOG(AREAUTIL)	-0.385831	0.014476	-26.65302	0.0000
QUARTOS	0.064349	0.007477	8.606508	0.0000
ARRUMOS	-0.032855	0.011417	-2.877674	0.0040
JARDIM	0.032432	0.009052	3.582719	0.0003
ISOLADA	0.027674	0.010905	2.537800	0.0112
NOVO	0.043288	0.010013	4.323179	0.0000
DTA_2006	0.077173	0.035570	2.169629	0.0301
DTA_2007	0.085979	0.030294	2.838167	0.0046
DTA_2008	0.108169	0.030461	3.551070	0.0004
DTA_2009	0.102824	0.030557	3.364958	0.0008
DTA_2010	0.101176	0.030420	3.325952	0.0009

R-squared	0.182497	Mean dependent var	6.816169
Adjusted R-squared	0.180031	S.D. dependent var	0.296114
S.E. of regression	0.268138	Akaike info criterion	0.208644
Sum squared resid	262.2119	Schwarz criterion	0.228994
Log likelihood	-369.7143	Hannan-Quinn criter.	0.215890
F-statistic	74.01312	Durbin-Watson stat	1.780317
Prob(F-statistic)	0.000000		

Table 14: Pooled regression for all periods (houses)

The annual dummies for 2004 and 2005 were not significant and therefore were excluded from the model. This inference is probably due to the scarcity of observations during the early years of the period (45 in 2003, one in 2004 and 41 in 2005).

Year	2006	2007	2008	2009	2010
RPPI	100	100.8845	103.1481	102.5983	102.4293
% Change	-	0.88%	2.24%	-0.53%	-0.16%

Table 15: RPPI for houses (time dummy method)

Given this limitation, we can only build an index from 2006 onward. The values can be seen in Table 15, and its evolution is plotted in Figure 11. The curve shows a positive trend between 2006 and 2008, followed by a slight decrease in 2009 and 2010. Given the limitations of this last estimation, we cannot build a global index that aggregates the last two indices calculated for the two sub-markets.

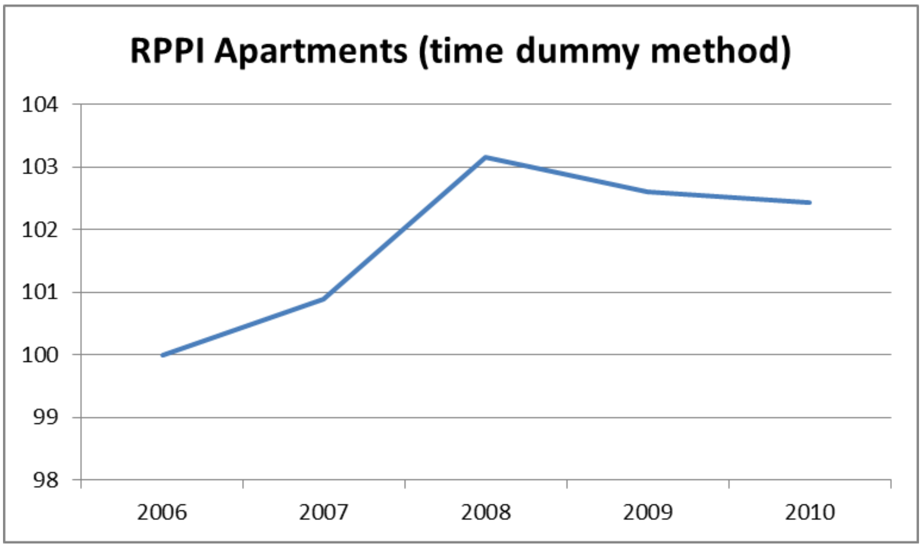

Figure 11: RPPI evolution (time dummy method): houses

3.2.2. Imputation Method

The time dummy variable method, as we referrer above, presents some problems that jeopardize the quality indices, namely with regard to the property of circularity. The imputation method overcomes this problem with hedonic yearly estimations and further construction of an annual chain index. The econometric model is similar to the above (Equation (7)) after removal of annual dummies and it is estimated for each sub-markets and for each year. We use for this exercise to logarithmic form just for simplification. The model takes the following form:

$$ln(p_n) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{nk} + u_n \qquad (8)$$

From the results of annual regressions, we calculate the price indexes using as the implicit quantities the average values of each of the characteristics of the dwelling in each year. Hereinafter, it is up to the user to opt for RPPI construction between Laspeyres, Paasche and Fisher type, or even build chain index, subsequently using the previous period as the base year.

The RPPI for the year 2004 compared to 2003, using the methodologies of Laspeyres and Paasche take the following expressions:

$$I_{Laspeyres}^{2004} = \frac{exp\left(\sum_{K=0}^{K} \widehat{\beta_K^{2004}} \overline{x_K^{2003}}\right)}{exp\left(\sum_{K=0}^{K} \widehat{\beta_K^{2003}} \overline{x_K^{2003}}\right)} \times 100 \qquad (9)$$

$$I_{Paasche}^{2004} = \frac{exp\left(\sum_{K=0}^{K} \widehat{\beta_K^{2004}} \overline{x_K^{2004}}\right)}{exp\left(\sum_{K=0}^{K} \widehat{\beta_K^{2003}} \overline{x_K^{2004}}\right)} \times 100 \qquad (10)$$

Where $\overline{x_K^I}$ corresponds to the mean value of characteristic, k, for the year i, $\widehat{\beta_K^I}$ to the estimated value of the parameter associated to the characteristic K, in year i and by definition$x_0^I = 1$.

Table 16 shows, for the apartments, the prices shadows obtained by annual regressions and the quantities obtained through the mean value of the sample characteristics of the property each year.

	Shadow prices						
Year	**β0**	**β1**	**β2**	**β3**	**β4**	**β5**	**β6**
2003	9.0772	-0.5331	0.0933	-0.0885	0.0564	0.0242	0.2037
2004	6.4369	0.1700	-0.2462	-0.1229	0.2809	0.7567	0.2660
2005	9.2795	-0.5464	0.0548	0.0058	-0.0554	0.0203	0.2535
2006	9.2173	-0.5261	0.0776	-0.0888	-0.0507	0.1354	0.2031
2007	9.1889	-0.5417	0.0869	-0.0471	0.0259	0.1056	0.2442
2008	9.1726	-0.5325	0.0654	-0.0267	0.0606	0.1221	0.1728
2009	8.9475	-0.4859	0.0682	-0.0193	0.0333	0.1989	0.1808
2010	8.9236	-0.4915	0.0633	-0.0312	0.0667	0.1575	0.2225
	Implicit quantities						
Year	**C**	**LOG(AREA)**	**QUARTOS**	**ARRUMOS**	**GARAGEM**	**NOVO**	**APU**
2003	1.0000	4.7525	2.2439	0.4329	0.6829	0.3537	0.8049
2004	1.0000	4.5703	2.0000	0.2941	0.4118	0.2353	0.5882
2005	1.0000	4.6305	2.0000	0.1693	0.2910	0.3650	0.8360
2006	1.0000	4.6614	2.1419	0.3135	0.5341	0.2154	0.8844
2007	1.0000	4.6668	2.1646	0.3323	0.6723	0.2840	0.9106
2008	1.0000	4.6364	2.1006	0.3070	0.6651	0.3075	0.9015
2009	1.0000	4.6766	2.1672	0.3345	0.6876	0.3921	0.9249
2010	1.0000	4.6543	2.1376	0.3606	0.7299	0.2709	0.9196

Table 16: Shadows prices and implicit quantities (Apartments, imputation method).

Year	Laspeyres Index	Paasche Index	Fisher Index	Chain Index
2003	100.0000	100.0000	100.0000	100.0000
2004	147.2566	120.3691	133.1358	147.2566
2005	105.7628	109.1434	107.4398	89.5228
2006	110.9383	111.1072	111.0227	106.4580
2007	112.1414	111.6960	111.9185	100.2270
2008	107.7678	106.5437	107.1540	95.1246
2009	110.0183	109.5129	109.7653	101.1710
2010	107.3265	105.8952	106.6084	98.0953

Table 17: RPPI indexes, imputation method (Apartment).

Table 17 shows the various indices (Laspeyres, Paasche, Fisher and chain). For the index Chain, we used Laspeyres weights, reporting systematically to the previous year. The chain indices resolve the issue of circularity index as they do not suffer from the same distortion effects observed for synthetic indices (Laspeyres and Paasche) due to non-homogeneity of the products compared between two different dates. Figure 12 plots the comparative evolution of the four RPPIs over the period. The first three (Laspeyres, Paasche and Fisher) evolve very similarly, except for the 2004 peak, where the Laspeyres index is more sensitive. The chain index coincides with the Laspeyres in 2004 taking into account the choice of the weighting. From 2005 onward, it stands out slightly from the other indexes with lower values. While synthetic indices (Laspeyres, Paasche and Fisher) present similar values, the chain index, despite evolving in the same direction, shows differences ranging from 2-5 percentage points below.

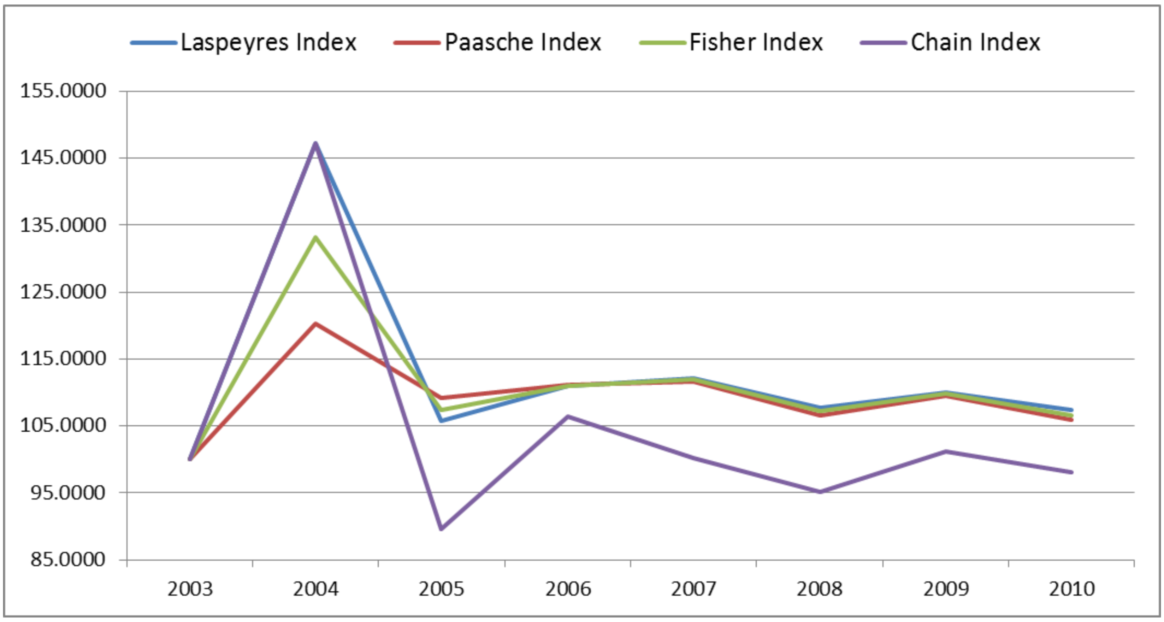

Figure 12: imputation RPPI (apartments)

Considering the scarcity of observations concerning the houses (villas), we do not reproduce the present exercise on these sub-markets. Thus, a comparative analysis will focus only at the level of the price index for apartments.

3.3. Comparative analysis

In this section, we proceed with the comparison of the various RPPIs calculated from the stratification method based on measures of central tendency and from the hedonic methods, with annual dummies variable and also using the imputation approach. We only compare the indexes concerning the apartment's sub-markets, leaving aside the house sub-market due to the observation scarcity. In the same graph Figure 13 we plot:

- Stratification Index based on weighted Fisher
- Imputation index with a weighting of Fisher
- Time dummy index

All index values as well as the respective annual rate of change can be seen in Table 18. The three indexes show the same trend, with similar values which tend to converge even further from 2005 onward. The imputation index seems more unstable, especially in the early period following the peak prices of 2004. The band values separating the three series varies from 10 percentage points in 2004, down to less than 1 percentage point in 2008 and 2009, and finally to about 1.5 percentage points in 2010.

Year	2003	2004	2005	2006	2007	2008	2009	2010
Imputation method	100.00	133.14	107.44	111.02	111.92	107.15	109.77	106.61
Imputation method (% change)	-	33.14%	-19.30%	3.33%	0.81%	-4.26%	2.44%	-2.88%
Stratification method	100.00	127.12	112.74	111.87	115.25	110.22	112.08	108.22
Stratification method (% change)	-	27.12%	-11.31%	-0.78%	3.02%	-4.36%	1.68%	-3.44%
Time dummy method	100.00	122.48	109.33	113.24	113.11	107.65	109.44	107.33
Time dummy method (% change)	-	22.48%	-10.73%	3.57%	-0.12%	-4.83%	1.67%	-1.93%

Table 18: RPPI, comparative analysis (apartments)

According to the literature (Eurostat, 2011), it is not possible to prescribe an universal method for all situations. The choice of the appropriate methodology may vary with the objectives pursued with the construction of the index, and also the availability of data in quantity and quality. For example, if the purpose is to build an index to assess the wealth or agents and its creditworthiness or solvability, the house sampling should be more comprehensive in order to be more representative of the existing housing stock, and, as such, should not comprise only the trade dwelling.

If, in contrast, the price index is designed to measure housing price dynamics or to be incorporated into a broader basket of prices designed to measure the evolution of the general price level, then the house sample estimation should be representative of the estate sold over a given period of time, thus favoring the flow over the stock measure. We have also seen that the stratification method is more susceptible to the bias effects, implying therefore more degrees of freedom, although it is less demanding from the point of view of quality and accuracy of information. In contrast, the hedonic methods are less demanding on the quantity of information (number of observations) provided that the most complete information about the most relevant characteristics of each property is guarantee.

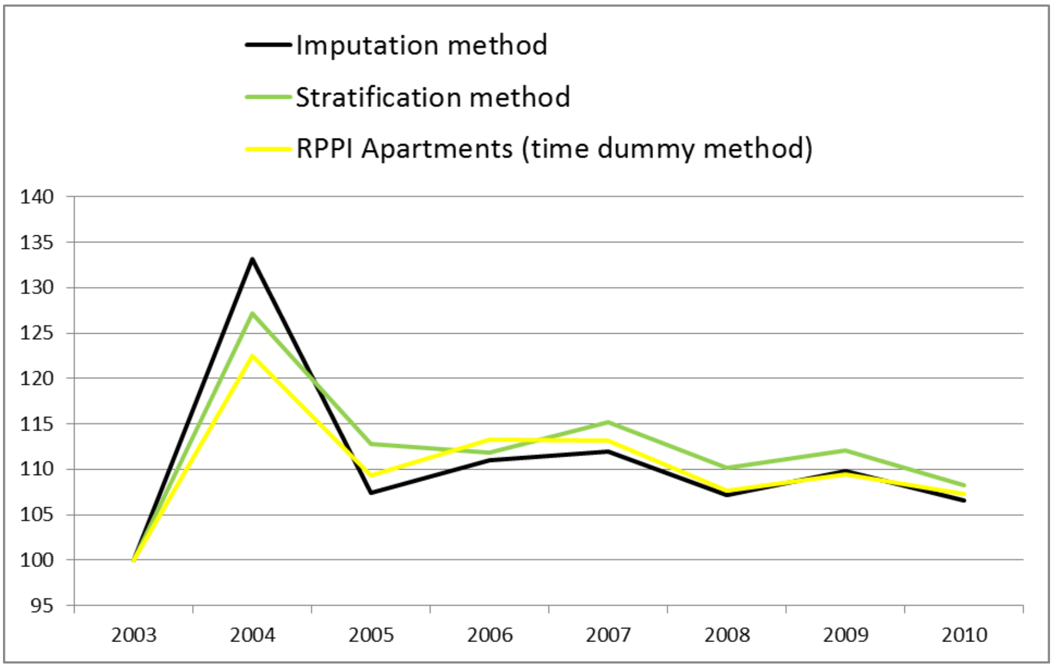

Figure 13: RPPI, comparative analysis (apartments)

4. Spatial data analysis

Hedonic models are normally based in cross-section data. In the present section, we intend to focus our attention on how the spatial effects affect the estimation of hedonic price. As we know the presence of autocorrelation and spatial heterogeneity profoundly affects the quality of the models, generating OLS estimators inconsistent and inefficient (Baumont, 2009). Housing prices are vulnerable to spatial effects for several reasons.[6] The first one concerns the accessibility (to CBD or any other centrality or equipment). The second has to do with the effects of neighborhood or how housing price capitalizes positive or negative externalities arising from urban dynamics. The third and final reason has to do with shared structural similarity to the scale of neighborhoods that are normally constructed as a whole and at the same time. As such, and considering the potential spatial autocorrelation, several criticisms may be pointed at these models, mostly related with the existence of multicollinearity, endogeneity, biasedeness and the existence of specification errors. These problems may seriously affect the robustness of the convergence coefficient and produce misleading outcomes (Anselin, 1988).

The hedonic equations incorporate in most cases a set of spatial variables designed to capture some of the special effects mentioned above. However, measurement of these spatial effects and the choice of appropriate variables, presents numerous and practical difficulties. According to (Anselin, 1988; LeSage, 2008) among others, the spatial data analysis with the introduction of the geographical dimension, namely in the presence of spatial autocorrelation, allows not only to capture the spatial effect, but also to improve the estimation and prevision since spatial dependence violates some of the Gauss-Markov assumptions of the OLS estimation (cross section observations are no longer independent) producing inefficient estimators. Two kinds of spatial effects are pointed out in the literature, namely: (i) spatial autocorrelation, revealing that contiguous regions may influence each other's performance through spillover effects and (ii) spatial heterogeneity, whenever the same functional form is erroneously considered for all regions (for comprehensive references about spatial econometric see for instance

[6] These spatial effect are well described in (Baumont, 2009).

(Anselin, 1988; Le Gallo, 2002; LeSage, 2008). Spatial autocorrelation, in turn, can be of two types: the spatial autoregressive dependence, in which the dependence is attached to contiguous economic variables and the spatial autocorrelation in the disturbance term, in which the spatial dependence is captured in the error term due to omitted variables or deficient functional form.

4.1. The analytical framework

For this exercise we will use the same database, however focusing our attention in the apartments, thus excluding house market. With this definition, we favor the quality of observations and also the homogeneity of the sample.

To capture the interdependence between locations, it is necessary to consider the relative positions of each location. For this, we must fix exogenously the topology of the space system, building a weight matrix (Le Gallo, 2002). This matrix is a square matrix, W, where each term, $w_{i,j}$ represents the way how the two location I and j connect to each other. The most widely type of matrix used is the contiguity matrix, in which the term $w_{i,j}$ equal to 1 if the two locations i and j are contiguous and zero otherwise. It is also possible to use Euclidian distances instead of a binary variable. The weights matrixes are usually standardized with each line sum equal to one.

The choice of spatial weights is a key issue in spatial model. In the present work we adopt a standard approach, built on geographical proximity (contiguity), considering however that geometric space is not always adequate to fully embrace the complexities of spatial structures (Bhattacharjee, Castro, & Marques, 2012; Marques, 2012). As such, we consider that a future and necessary extension of the present thesis will consist of developing an alternative methodology in order to propose a framework to capture and analyze the intangibility of space, both in terms of spatial heterogeneity and spatial spillovers. In this potential multidimensional non-Euclidean space, which does not necessarily conform to the usual physical concept of distance or contiguity, the objective is to use a more complex representation of space exploring the economic and social connections that link our geographical units.

In our sample, the observations are not classified by any geographic information system, i.e., we do not have the coordinates of each individual residence. To circumvent this limitation we have built our contiguity matrix using the same methodology as in (Marques, 2012), delimiting 75 different geographical areas. These zones correspond to homogenous territories, normally smaller than parishes, and represent neighborhoods, centers or other clusters which similar pattern (see Figure 14).

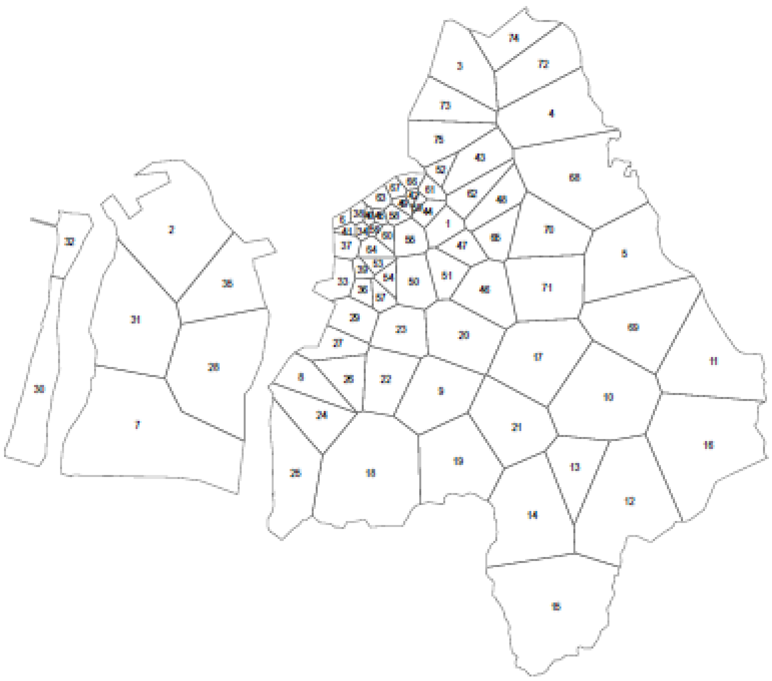

Figure 14: Codes of the 75 zones used in spatial weight matrix. Source: (Marques, 2012).

Thus the criterion for adjacency passes by assigning the value 1 to the term associated with a specific flat if the same flat is located within the zone or its influence area (contiguous), and zero otherwise. As mentioned before, the final matrix is standardized so that the elements of a row sum up to one. The working spatial contiguity matrix, W, appears in Figure 15. As we can see from the figure, the matrix W shows a sparse structure with most of the non-zero elements residing near the diagonal.

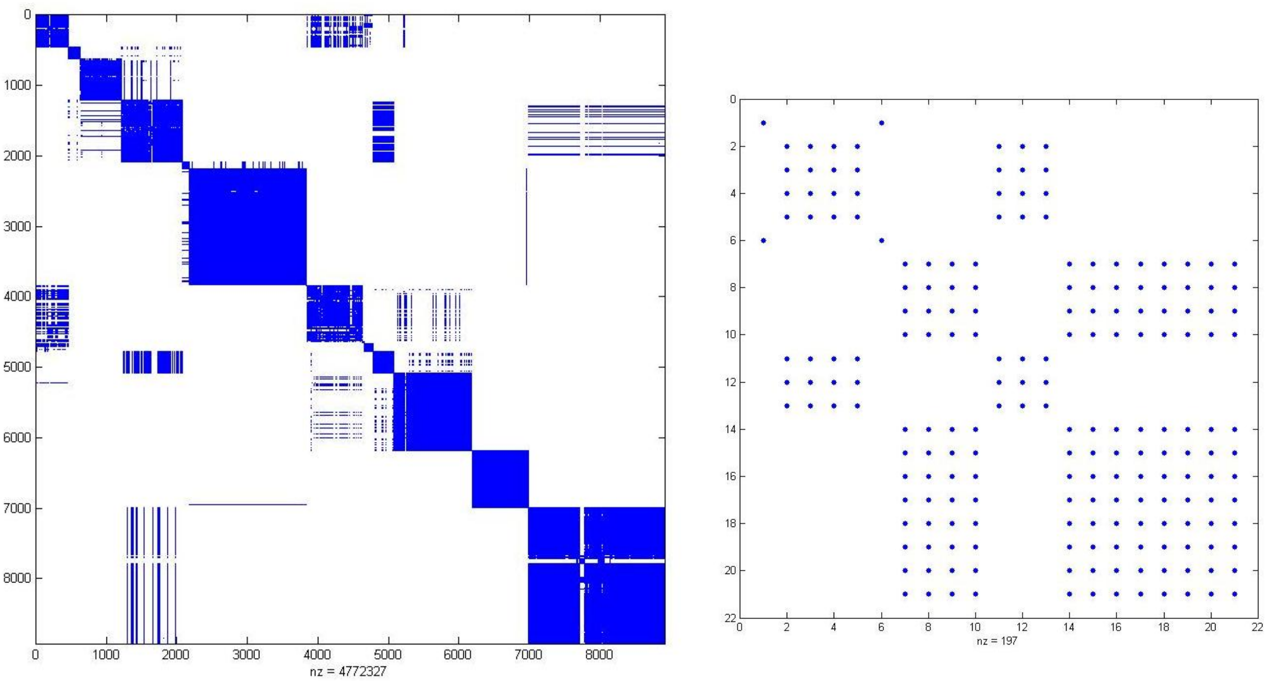

Figure 15: full 8921×8921 spatial weights (contiguity) matrix (right side), and a partial view (left side).

First we measure the degree of spatial autocorrelation of log prices. The spatial autocorrelation is based on the Moran's statistic (Moran's I), which can be represented by the expression:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j} \qquad \textbf{(11)}$$

in which ϖ_{ij} represents the $\{i, j\}$ element of the spatial contiguity matrix, W, such as $\varpi_{ij} = 1$ if apartments i and j are neighbours and $\varpi_{ij} = 0$ otherwise, x_{it} represents the logarithm of price per square meter of apartment i, and n corresponds to the number of observations. Moran's I estimates the linear dependence between a variable in a specific location and the mean of the same variable in the neighborhood. Note that Moran ‘I is a global measure of spatial autocorrelation. As such it does not describe any local pattern of spatial autocorrelation, like clusters of high or low prices houses.

A positive value of Moran ‘I statistic indicates the presence of correlation between the dwelling price and the average price of the surrounding habitations. The Moran scatter plot depicts the log-price on the horizontal axis with the average values of the neighbouring houses on the vertical axis. The four quadrants in the scatter plot show, respectively, (i) the houses with high prices associated with neighbouring houses also with high prices (red points), (ii) the houses with low prices associated with neighbouring houses also with low prices (cyan points), (iii) the houses with low prices associated with neighbouring houses with high prices (green points), (iv) the houses with high prices associated with neighbouring houses with low prices (purple points). A positive value of I implies that most houses remains in the first and second quadrants. The remains quadrants are considered as atypical locations.

The next step consists in detecting the presence of spatial effects through several tests. These testes are crucial in order to adopt the correct model specification to fully integrate the spatial effects. We use the same model built to estimate the hedonic regression related to apartments in the above sections. Among the different specifications we consider a model in which the spatial dependence is associated to the lagged dependent variable (equation 12); the spatial error model (equation 13), in which only disturbances exhibit spatial dependence; and finally the general version of the spatial model (equation 14), including both the spatial lagged term as well as the spatial correlated error term. W corresponds to the spatial contiguity matrix defined above and is common to all three specifications and ε an independent and identically distributed process.

$$ln(p_n) = \beta_0 + \rho W ln(p_n) + \sum_{\tau=1}^{T} \delta^\tau D_n^\tau + \sum_{k=1}^{K} \beta_k x_{nk} + \varepsilon_n \qquad (12)$$

$$ln(p_n) = \beta_0 + \sum_{\tau=1}^{T} \delta^\tau D_n^\tau + \sum_{k=1}^{K} \beta_k x_{nk} + \boldsymbol{u}_n \quad with\ u_n = \Lambda W u_n + \varepsilon_n \qquad (13)$$

$$ln(p_n) = \beta_0 + \rho W ln(p_n) + \sum_{\tau=1}^{T} \delta^\tau D_n^\tau + \sum_{k=1}^{K} \beta_k x_{nk} + u_n \qquad (14)$$

1.1 The exploratory spatial data analysis: results and discussion

For this exercise, we will use the same database. The lack of capacity of computer memory prevents us however to use all the observations. Thus we will use only the data from 2005 to 2009, comprehending a total of 8162 observations. All estimations are carried out in Matlab using the general maximum likelihood method.7

We start by measuring the global autocorrelation of the dependent variable, i.e., the logarithm of prices through the Moran 'I Statistic. As it can be seen in Table 19, prices are positively auto correlated in space, with a Moran 'I of 0.49 highly significant. This means that dwellings tend to group themselves according to price, with the segments more expensive concentrated in certain areas and the lower segments in other different areas.

Variable	Moran's I	Moran's I stat.	P value
Log (price	0.4916	45.85	0.0000

Table 19: Moran's I statistic for the log of price per meter.

The Moran scatter plot (Figure 16), in which the average value of neighboring dwellings are plotted against the value of central dwelling, shows, as expected, that most of observation are located in the quadrant HH and LL. In particular we have 2717 observation in the HH quadrant, 841 in the LH quadrant, 3323 in the LL quadrant and 1281 in the HL quadrant, which means that 74% of apartments are located in the HH and LL quadrant. The remaining observations (2122) are located in the atypical quadrants. The Moran's I Statistic give us valuable indications on house pricing to concentrate and form clusters (Arbia, 2001). However, it tells us nothing about the spatial location of these specific manifestations of agglomeration. Thus, these global indexes, if relevant, can be an invitation to explore other local measures of agglomeration as statistical LISA (Local

[7] All calculation are based on (Elhorst, 2003) and (LeSage, 1999). We use the authors Toolbox function available respectively at http://www.regroningen.nl/elhorst/software.shtml and http://www.econ.utoledo.edu.

Indicator of Spatial Association) which decomposes the Moran's I Statistic in order to identify the individual contribution of each local site (in our case, each municipality). This is certainly a matter for future research.

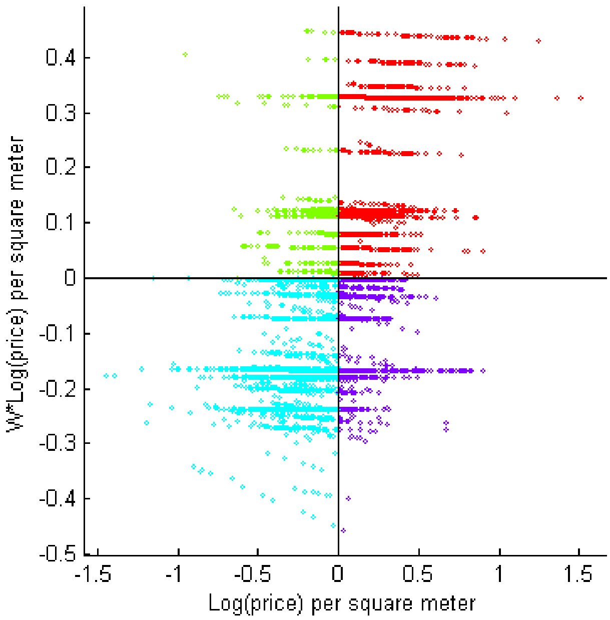

Figure 16: Moran scatterplot for log(price)

Finally we use a spatial econometric methodology to estimate the time dummy hedonic price model (Equation (7)). First, as a starting point we estimate the general pool OLS modelhe next step consists in detecting the presence and type of spatial effects, in order to evaluate whether the spatial lag model or the spatial error model is the most appropriate to describe the data. We follow the robust LM-tests described in (Elhorst, 2003) which test the type of spatial dependence based on the residual of the non-spatial model.[8]

[8] Matlab routines available at www.regroningen.nl/elhorst.

The results exhibited in the first column of Table 20, indicate the presence of spatial dependence both in the lagged dependent variable and in the residual. As such, all specifications must be tested: the mixed autoregressive model in which the spatial dependence in associated with the lagged dependent variable (SAR model described in equation 12); the spatial error model (SEM model described in equation 13), in which only disturbances exhibit spatial dependence; and finally the general version of the spatial model (SAC model described in equation 14). Both the SAR and SEM specification make sense. In the former, the spatial dependence in the log(price) variable expresses the fact that the price of a house is affected by its neighborhood. In the latter the spatial dependence is associated with a statistical nuisance which may occur from various forms of misspecification (omitted variables, the lack of adequate neighborhood measures etc.) (Baumont, 2009).

Considering that the non-consideration of spatial effects may compromise the quality of the estimators, the adoption of models that include spatial interaction is clearly justified. Comparing the three spatial models (SAR, SEM and SAC), and considering the results of the spatial autocorrelation test of the OLS residual, we concluded that the general special model (SAC) given by equation (14), is adequate for the dummy variable hedonic model. Also, the SAC R^2 is higher compare to the remaining specification. This means that the model variable, including the location one, do not allow to capture the spatial variations of the housing price distribution. Comparing the OLS and the SAC specification, all coefficients remain significant. However, there are substantial differences in their values. For instance, a garage increases by 1.7 percent the per square meter price of the apartment with the OLS model while it increases by 5 percent according to the SAC estimation. For the OLS model, a new apartment is worth more 22.1 percent in relation to an apartment not new against 18 percent according to the SAC estimation). Finally the APU location increases the price about 21.5 percent for the OLS estimation against only 9.1 percent for the SAC estimation.

Models	OLS	SAR	SEM	SAC
Estimation				
Obs	8162	8162	8162	8162
R2	0.4413	0.4165	0.6678	0.6781
Const.	9.1118 (p=0.0000)	6.625161 (p=0.0000)	9.152611 (p=0.0000)	8.97602 (p=0.0000)
log(area)	-0.5487 (p=0.0006)	-0.529136 (p=0.0000)	-0.536451 (p=0.0000)	-0.536403 (p=0.0000)
Rooms	0.0902 (p=0.0000)	0.093106 (p=0.0000)	0.097124 (p=0.0000)	0.097313 (p=0.0000)
Garage	0.0174 (p=0.0000)	0.026518 (p=0.0000)	0.049785 (p=0.0000)	0.049609 (p=0.0000)
New	0.2214 (p=0.0000)	0.195518 (p=0.0000)	0.179841 (p=0.0000)	0.179621 (p=0.0000)
APU	0.2146 (p=0.0000)	0.132109 (p=0.0000)	0.097469 (p=0.0000)	0.091043 (p=0.0002)
Y2006	-0.0139 (p=0.0000)	-0.018385 (p=0.0272)	-0.027503 (p=0.0002)	-0.02737 (p=0.0002)
Y2007	0.0035 (p=0.0233)	-0.00022 (p=0.0974)	-0.006666 (p=0.0897)	-0.006298 (p=0.0726)
Y2008	-0.0078 (p=0.0024)	-0.014006 (p=0.0884)	-0.030646 (p=0.0000)	-0.030303 (p=0.0000)
Y2009	-0.0117 (p=0.0063)	-0.027407 (p=0.0020)	-0.062385 (p=0.0000)	-0.062086 (p=0.0000)
ρ		0.350949 (p=0.0000)		0.027597 (p=0.0000)
ʎ			0.9000 (p=0.0000)	0.8980 (p=0.0000)
Autocorrelation tests				
Moran's I	0.9878 (p=0.0000)	-	-	-
LM lag (robust)	63.1163 (0.0000)	-	-	-
LM error (robust)	69.5428 (0.0000)	-	-	-

Table 20: Estimation results and spatial tests (Source: own calculations based on Sapo-casa database)

Concerning the RPPIs constructed through the time dummy variable method, we note as we did for the other variables the same trends between the OLS model and the SAC model, although with important differences in the level of estimators (Table 20). In terms of the two series of price indexes, although both point to a fall in prices since 2007 (see Figure 17), we find that this fall is far more pronounced in the series based on the SAC model with spatial dependence. Indeed, as you can see in Table 21, the difference between the two series reaches a maximum of nearly three percentage points, with the series OLS pointing in 2009, to a price reduction of around 0.39 percent vs. 3.13 percent less for the series with the SAC model. Thus, the presence of spatial interaction, which presence has been detected, affects significantly the value of the estimators, with a direct impact on the calculation of price indices.

Year	2005	2006	2007	2008	2009
OLS	100.00	98.62	100.35	99.22	98.84
SAC	100.00	97.30	99.37	97.02	93.98
% Change (OLS)	-	-1.38%	1.76%	-1.12%	-0.39%
% Change (SAC)	-	-2.70%	2.13%	-2.37%	-3.13%

Table 21: RPPIs evolution, with (SAC) and without spatial effects (OLS).

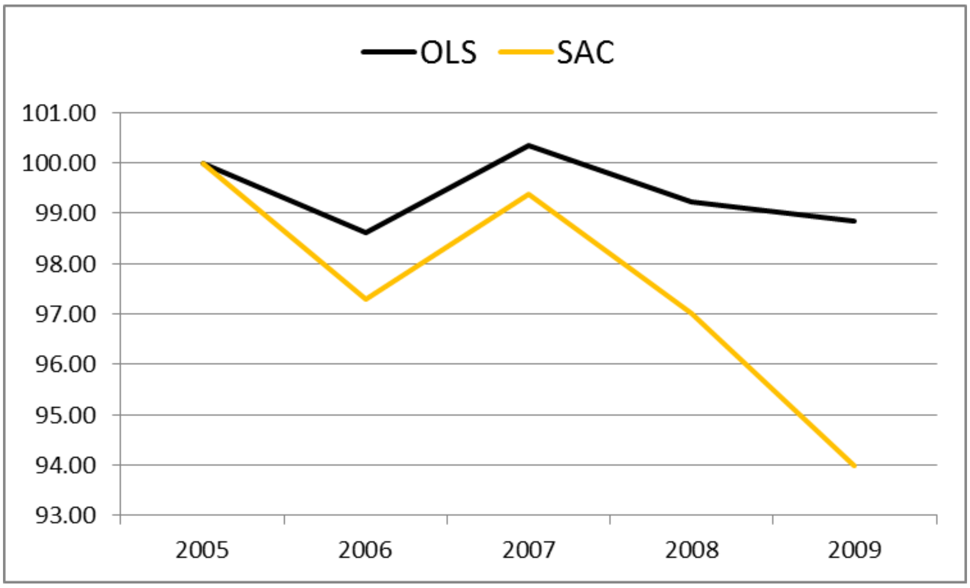

Figure 17: RPPIs graph with (SAC) and without spatial effects (OLS).

5. Conclusion

Throughout this paper, we test a set of methodologies for the construction of residential property price indexes, in line with the latest developments in the literature. From a database of “Sapo-casa”, we obtained a set of results that showed similar trends but not necessarily identical in level. These results can be considered consistent with the empirical reality whose standard, has been given by the INE (National Statistics Institute) "Survey on bank evaluation on housing".

Although the methodologies used have focused on the same database, it is clear that the quantity and quality of available data are decisive for the choice of the type of price index to build. The methods, based on the first moments of the price distribution (or median), are very attractive for its simplicity and allow, provided there is sufficient degrees of freedom, to obtain disaggregated prices for various sub-markets, therefore, revealing good and precise information to the market. When the number of observations is scarce, hedonic methods represent a valid alternative, provided that the available observations are accompanied by qualitative information about the characteristics of the property traded.

In line with the literature, home prices show a high degree of spatial correlation not captured by traditional models. In this sense, the results from spatial econometric models show significant differences when compared with the results obtained with the traditional OLS model, thus raising relevant questions about the consistency of these estimators. Even in full hedonic models incorporating a wider set of variables intended to capture the territorial spatial effects, the literature reveals frequently the presence of auto correlation in the residuals (Baumont, 2009). This finding should have, in our opinion, a significant impact in the construction of a methodological framework for the development of residential property price indexes with national coverage, even considering the necessity of these methods being parsimonious and replicable in different territories. Given the difficulty of harmonizing the development of territorial variables impacting on house prices (CBO, services, equipment, infrastructure etc.) it seems that the tools of spatial econometrics will be an instrument to take increasingly into account in this matter.

Still considering the contribution to a methodological framework for the construction of real estate price indices or residential property price indexes, it is clear the need to find, upstream, the best mechanisms for the creation of a database that feeds the estimation of indexes. In this sense, the literature and the various practices point to four data sources: the notarial sources for transactions registrations, the fiscal institutions that assess the assets for tax purposes, the real estate agencies that promote business and finally the banking that provides credit for the purchase of housing. Any of these solutions have advantages and disadvantages. Considering the need to monitor the evolution of the real estate price level, eventually integrated in a wider framework of building a broader indicator of prices evolution, the crucial observations consist fundamentally of the trade register if possible performed in real time. In this case, the recording notary shall certainly be the more comprehensive mechanism, in that not all credit request lead to sales and many of these transactions are performed outside the domain of real estate agencies or without the bank intermediation. If the objective is to assess property wealth for national accounting or fiscal purpose or even to assess solvency of institutions or agents, the databases of national or regional fiscal institutions are certainly a source unavoidable.

Whatever the solutions or protocol to adopt in order to operationalize the constitution and updating of a real estate database, the involvement of public institutions, either through the INE, or the academic institutions, seems of utmost importance in a context where new technological solutions related to georeferencing data open new perspectives to econometricians. Therefore, and taking into account the results that reinforce the importance of spatial effects in the distribution of prices, it is no longer possible to think about the creation and updating of real estate databases without including the issue of developing, in parallel, a geographical information system that allows to incorporate the spatial component in econometric models routines based on georeferenced data.

References:

Anselin, L. (1988). *Spatial econometrics: methods and models* (Vol. 4): Springer.

Arbia, G. (2001). The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical systems, 3*(3), 271-281.

Balk, B. M. (1995). Axiomatic price index theory: a survey. *International Statistical Review/Revue Internationale de Statistique*, 69-93.

Basu, S., & Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. *The Journal of Real Estate Finance and Economics, 17*(1), 61-85.

Baumont, C. (2009). Spatial effects of urban public policies on housing values*. *Papers in Regional Science, 88*(2), 301-326.

Bhattacharjee, A., Castro, E., & Marques, J. (2012). Spatial interactions in hedonic pricing models: the urban housing market of Aveiro, Portugal. *Spatial Economic Analysis, 7*(1), 133-167.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252.

Cropper, M. L., Deck, L. B., & McConnell, K. E. (1988). On the choice of funtional form for hedonic price functions. *The Review of Economics and Statistics*, 668-675.

Dubujet, F. (2000). Les indices de prix Notaires-Insee. *Unpublished manuscript, 23*, 2000.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International regional science review, 26*(3), 244-268.

Eurostat. (2011). *RPPI Handbook*: Eurostat.

Fleming, M. C., & Nellis, J. (1984). *The Halifax house price index: technical details*: Halifax Building Society.

Goldfarb, B., & Pardoux, C. (2011). *Introduction à la méthode statistique-6e édition-Économie, gestion: Économie, gestion*: Dunod.

Griliches, Z. (1971). Hedonic price indexes for automobiles: an econometric analysis of quality change. *Price indexes and quality change*, 55-87.

Hill, P. (1990). RÉCENTS DÉVELOPPEMENTS DE LA THÉORIE ET DE LA PRATIQUE DES INDICES. *OCDE org*.

Hill, R. J. (2004). Constructing price indexes across space and time: the case of the European Union. *American Economic Review*, 1379-1410.

Hill, R. J., & Melser, D. (2008). Hedonic imputation and the price index problem: an application to housing. *Economic Inquiry, 46*(4), 593-609.

Hyde, S. (1999). Likelihood based inference on the Box-Cox family of transformations: SAS and MATLAB programs. *Dept of Mathematical Sciences, Montana State University, Billings 34p*.

Lancaster, K. J. (1966). A new approach to consumer theory. *The Journal of Political Economy, 74*(2), 132-157.

Le Gallo, J. (2002). Econométrie spatiale: l'autocorrélation spatiale dans les modèles de régression linéaire. *Economie & prévision*(4), 139-157.

LeSage, J. P. (1999). *Spatial econometrics*: Regional Research Institute, West Virginia University.

LeSage, J. P. (2008). An introduction to spatial econometrics. *Revue d'économie industrielle*(3), 19-44.

Maleyre, I. (1997). L'approche hédonique des marchés immobiliers. *Etudes foncières*(76), 22-29.

Mark, J. H., & Goldberg, M. A. (1984). Alternative housing price indices: an evaluation. *Real Estate Economics, 12*(1), 30-49.

Marques, J. J. L. (2012). The notion of space in urban housing markets. *Ph.D. Thesis. University of Aveiro: Portugal*.

McDonald, C., & Smith, M. (2009). *Developing stratified housing price measures for New Zealand*: Reserve Bank of New Zealand.

Osborne, J. W. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation, 15*(12), 1-9.

Prasad, N., & Richards, A. (2006). *Measuring housing price growth–using stratification to improve median-based measures*: Reserve Bank of Australia.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy, 82*(1), 34-55.

Shiller, R. J. (1991). Arithmetic repeat sales price estimators. *Journal of Housing Economics, 1*(1), 110-126.

Taylor, L. O. (2003). The hedonic method. In *A primer on nonmarket valuation* (pp. 331-393): Springer.

Wang, F. T., & Zorn, P. M. (1997). Estimating house price growth with repeat sales data: what's the aim of the game? *Journal of Housing Economics, 6*(2), 93-118.

Waugh, F. V. (1928). Quality factors influencing vegetable prices. *Journal of farm economics, 10*(2), 185-196.